

No. d'identification

//////////

**UNIVERSITÉ MONTPELLIER II**

**U.F.R SCIENCES ET TECHNIQUES DU LANGUEDOC**

**THESE**

pour obtenir le grade de

**DOCTEUR EN SCIENCES**

**DE L'UNIVERSITÉ DE MONTPELLIER II**

**Ecole Doctorale : Biologie Intégrative**

**Formation Doctorale : Biologie de l'Evolution et Ecologie**

**Section CNU : Biologie des Populations et Ecologie**

présentée par

**Cécile GRENIER**

**DIVERSITÉ GÉNÉTIQUE DES SORGHOS CULTIVÉS  
GESTION DES RESSOURCES GÉNÉTIQUES ET CONTRIBUTION DES  
MARQUEURS MOLÉCULAIRES À LA CONSTITUTION DE CORE  
COLLECTIONS**

Soutenue le 4 février 2000 devant la commission d'examen composée de :

<b>Bernard DELAY</b> , Directeur de Recherches CNRS, Université Montpellier II	Président
<b>Gilles CHARMET</b> , Directeur de Recherches INRA	Rapporteur
<b>Grégoire THOMAS</b> , Professeur Université de Rennes	Rapporteur
<b>André CHARRIER</b> , Professeur ENSAM	Examineur
<b>Jean Christophe GLASZMANN</b> , Chef de Programme BIOTROP, CIRAD	Examineur
<b>Perla HAMON</b> , Maître de Conférences, Université Montpellier III	Directrice de thèse
<b>Marianne LEFORT</b> , Directrice de Recherches INRA Paris	Examinatrice
<b>Rodomiro ORTIZ</b> , Directeur de Programme GREP, ICRISAT	Examineur

No. d'identification

/ / / / / / / / / /

**UNIVERSITÉ MONTPELLIER II**  
**U.F.R SCIENCES ET TECHNIQUES DU LANGUEDOC**

**THESE**

pour obtenir le grade de

**DOCTEUR EN SCIENCES**  
**DE L'UNIVERSITÉ DE MONTPELLIER II**  
Ecole Doctorale : Biologie Intégrative  
Formation Doctorale : Biologie de l'Evolution et Ecologie  
Section CNU : Biologie des Populations et Ecologie

présentée par

**Cécile GRENIER**

**DIVERSITÉ GÉNÉTIQUE DES SORGHOS CULTIVÉS**  
**GESTION DES RESSOURCES GÉNÉTIQUES ET CONTRIBUTION DES**  
**MARQUEURS MOLÉCULAIRES À LA CONSTITUTION DE CORE**  
**COLLECTIONS**

Soutenue le 4 février 2000 devant la commission d'examen composée de :

<b>Bernard DELAY</b> , Directeur de Recherches CNRS, Université Montpellier II	Président
<b>Gilles CHARMET</b> , Directeur de Recherches INRA	Rapporteur
<b>Grégoire THOMAS</b> , Professeur Université de Rennes	Rapporteur
<b>André CHARRIER</b> , Professeur ENSAM	Examineur
<b>Jean Christophe GLASZMANN</b> , Chef de Programme BIOTROP, CIRAD	Examineur
<b>Perla HAMON</b> , Maître de Conférences, Université Montpellier III	Directrice de thèse
<b>Marianne LEFORT</b> , Directrice de Recherches INRA Paris	Examinatrice
<b>Rodomiro ORTIZ</b> , Directeur de Programme GREP, ICRISAT	Examineur

## REMERCIEMENTS

Tout d'abord, je tiens à remercier très chaleureusement ma directrice de thèse, Perla Hamon, pour sa confiance, son encouragement, sa rigueur, ses conseils. Je tiens à lui témoigner ma très grande reconnaissance pour son encadrement, effectué de main de maître.

Je remercie sincèrement Gilles Charmet et Grégoire Thomas de l'intérêt témoigné pour cette thèse, en ayant accepté d'en être les rapporteurs.

Je tiens à remercier vivement Bernard Delay, d'avoir présidé le jury de thèse, André Charrier, Jean Christophe Glaszmann, Marianne Lefort et Rodomiro Ortiz d'avoir bien voulu répondre positivement pour faire partie de ce jury.

Cette thèse s'est déroulée en grande partie loin de l'hexagone...

Merci à Paula Bramel-Cox pour m'avoir intégrée dans son équipe des Ressources Génétiques, à l'ICRISAT en Inde. Je tiens à remercier KE Prasada Rao pour sa grande patience lors de mon apprentissage de la taxonomie du sorgho, et Ashok Kumar qui m'a assistée dans toute les démarches administratives. Un grand *shukria* à Kamala et à Radha qui m'ont soutenue au cours de mon expatriation.

Puis il y a eu une étape aux Etats Unis, où j'ai été très chaleureusement accueillie dans le laboratoire des Ressources Génétiques des Plantes de l'USDA. Je remercie vivement Steve Kresovich pour avoir tout mis en oeuvre pour que tout se passe au mieux. Merci à toute son équipe. Merci à Marcio et à Gabriel.

Enfin, il y a eu un retour en France, et là encore, la liste des remerciements ne pourra être complète. Je remercie tout d'abord Jean-Christophe Glaszmann, qui m'a permis de bénéficier d'un accueil chaleureux à mon retour des tropiques - au mois de janvier - dans son laboratoire Biotrop. Je tiens à exprimer ma très grande gratitude envers Monique Deu, pour son aide, son soutien, ses encouragements quotidiens, ses conseils, nos discussions. Un très grand merci à Carole, qui a su transformer les moments difficiles en anecdotes inoubliables. Je tiens à remercier Jacques Chantereau pour son enthousiasme au transfert de connaissances. Merci également à Laurent Grivet pour les discussions très fructueuses. Un grand merci pour Brigitte Giudicelli, et pour Monique Costes pour leurs encouragements très chaleureux. A toute l'équipe de Biotrop et des ressources génétiques, je voudrais également témoigner mes vifs remerciements pour leur grande amitié, ainsi qu'à Fanny, Isabelle, Dominique...

Puis il y a ceux qui sont arrivés juste avant la thèse et qui ont continué à me témoigner leur amitié et soutien malgré mon enfouissement dans la rédaction... Un très grand merci à la petite Sophie pour ses puissants encouragements. Un grand merci à Léa, Pierre, Anne-Hélène, Claire...

Enfin, il y a ceux qui m'encouragent depuis de longue date, et à qui je dois beaucoup. Ils m'ont toujours très fortement soutenue, et plus particulièrement durant cette thèse, à chaque moment, et dans toutes les situations... Merci à mes parents et frères, présents tout au long de ces trois années, quotidiennement, et sans qui... ce qui suit n'aurait pu être...

# SOMMAIRE

INTRODUCTION	1
<b>A - LES RESSOURCES GÉNÉTIQUES</b>	<b>2</b>
I - Définition des ressources génétiques	2
II - Collecte des ressources génétiques	3
a) <i>espèces d'importance alimentaire</i>	3
b) <i>espèces cultivées à intérêt économique pour l'agriculture des pays tropicaux</i>	4
c) <i>les espèces sauvages</i>	5
III - Gestion des ressources génétiques	6
1) conservation <i>ex-situ</i>	7
2) conservation <i>in-situ</i>	8
3) conservation on-farm	8
IV - Evaluation des ressources génétiques	9
V - Core collection	10
1 - Définition	10
2 - Stratégies de constitution de core collections	11
2 - 1. Stratégies aléatoires	11
2 - 1. 1. <i>structuration de la collection de base</i>	11
2 - 1. 2. <i>échantillonnage</i>	12
2 - 2. Stratégies non-aléatoires	13
2 - 2. 1. <i>critères phénotypiques</i>	13
2 - 2. 2. <i>critères biochimiques ou moléculaires</i>	14
2 - 2. 3. <i>autres critères</i>	14
3 - Utilisation des core collections	15
<b>B - LE SORGHO</b>	<b>17</b>
I - La production en sorgho	17
II - Taxonomie	18
III - Origine géographique et domestication	19
IV - Les collections de sorgho	20
1 - La collection de l'ICRISAT	20
2 - La collection aux USA	21
<b>C - OBJECTIFS DU TRAVAIL</b>	<b>22</b>



<b>1<sup>ERE</sup> PARTIE - ETUDE DE LA COLLECTION DE BASE</b>	<b>24</b>
<i>CHAPITRE 1 : Diversité morpho-agronomique</i>	<i>27</i>
<i>CHAPITRE 2 : Organisation de la diversité</i>	<i>34</i>
<b>2<sup>EME</sup> PARTIE - CORE COLLECTION</b>	<b>39</b>
<i>CHAPITRE 1 : Stratégies de prélèvement aléatoire</i>	<i>40</i>
I - Structure de la collection de base	40
II - Comparaison entre trois stratégies d'échantillonnage	42
<i>CHAPITRE 2 : Impact des stratégies d'échantillonnage aléatoire vs. non aléatoire sur la diversité morpho-agronomique</i>	<i>45</i>
<b>3<sup>EME</sup> PARTIE - IMPACT DE DIFFÉRENTS TYPES DE PRÉLÈVEMENT SUR LA DIVERSITÉ NON ÉCHANTILLONNÉE</b>	<b>51</b>
<i>CHAPITRE 1 : Impact des stratégies d'échantillonnage aléatoire vs. non aléatoire sur la diversité moléculaire</i>	<i>52</i>
<i>CHAPITRE 2 : Recherche d'associations entre marqueurs</i>	<i>55</i>
<i>CHAPITRE 3 : Impact sur la diversité morpho-agronomique d'un échantillonnage basé sur la diversité moléculaire</i>	<i>58</i>
<b>CONCLUSION - DISCUSSION GENERALE</b>	<b>64</b>
<b>RÉFÉRENCES BIBLIOGRAPHIQUES</b>	<b>70</b>
<b>ANNEXES</b>	
<b>PUBLICATIONS</b>	

## INTRODUCTION

## INTRODUCTION

La gestion des collections de grande taille pose un problème majeur pour les centres de recherche responsables de la maintenance de la richesse génétique de telles espèces. L'Institut International de Recherche pour les Cultures des régions Tropicales Semi-Arides (ICRISAT), dépositaire de collections de très grands effectifs, doit assurer la maintenance pour un nombre considérable d'accessions (113 170 accessions maintenues dans la banque de semences en 1998). Le budget alloué au système de conservation à long terme, c'est-à-dire à -20°C, est très élevé. De plus, si les caractérisations, réalisées lors de l'entrée des variétés dans la collection, sont à peu près complètes sur l'ensemble des accessions, les étapes d'évaluation du matériel maintenu sont plus difficilement réalisables et plus coûteuses. Ainsi, le manque d'informations sur la collection ne peut permettre de bénéficier pleinement de la richesse des ressources génétiques.

Un travail en collaboration entre l'ICRISAT, le Centre International de Recherche Agronomique pour le Développement (CIRAD) et l'Institut de Recherche pour le Développement (IRD), auquel s'est ajouté le Département de l'Agriculture des USA (USDA), a été initié en décembre 1996. Il avait pour objectif d'évaluer la diversité génétique de collection de sorghos cultivés maintenue au Centre Asiatique de l'ICRISAT (IAC) et d'appliquer différentes stratégies pour la constitution de core collections. Ces ensembles de taille réduite devraient permettre de faciliter la gestion de cette collection et d'optimiser l'utilisation de la diversité génétique des sorghos cultivés.

## A - LES RESSOURCES GÉNÉTIQUES

### I - Définition des ressources génétiques

A partir des années 30, faisant suite aux travaux de Vavilov, la notion de ressources génétiques, limitées alors aux seules plantes cultivées, évolue vers une notion plus large. Les ressources génétiques concernent non seulement les espèces cultivées et les espèces sauvages -parfois les genres - plus ou moins apparentées, mais aussi les espèces sauvages présentant un intérêt direct ou indirect. Cependant, cette notion reste à l'état de théorie jusqu'à ces trois ou quatre dernières décennies. Dans les années 60, la prise de conscience de l'importance effective des ressources génétiques fait suite à une constatation. Depuis le développement industriel, la disparition de certaines espèces vivantes (règne végétal mais aussi règne animal) n'a cessé de s'accélérer. Les éléments écologiques et les caractéristiques démographiques sont très probablement des facteurs dominants dans la survie à court terme des populations et des espèces (Hamrick et Godt, 1995). De cette érosion de la diversité biologique est née la problématique liée aux ressources génétiques, avec en premier lieu une urgence majeure, celle de les collecter et de les conserver.

Aujourd'hui, les ressources génétiques sont devenues un enjeu majeur, impliquant des prises de décisions politiques, scientifiques et économiques (O'Brien, 1994). Cette évolution des mentalités n'est pas sans engendrer des difficultés. Actuellement, il faut faire face à une situation très paradoxale. D'un côté, tout concourt au développement d'une agriculture à haute productivité, utilisant des variétés améliorées présentant des rendements élevés, résistantes aux stress biotiques et abiotiques, adaptées à la mécanisation de l'agriculture..., **mais** ayant une base génétique peu diversifiée. De l'autre côté, est présent le souci de maintenir des variétés traditionnelles, source de grande diversité génétique, mais moins productives.

Face à l'érosion génétique et pour assurer une utilisation optimale des ressources phytogénétiques, les scientifiques se sont fixés deux objectifs principaux : 1) collecter les ressources génétiques et en décrire la diversité génétique ; 2) trouver des méthodes efficaces de conservation et d'utilisation des ressources génétiques, de façon à augmenter la base génétique des variétés améliorées.

### II - Collecte des ressources génétiques

La première priorité définie pour la collecte des ressources génétiques concerne des espèces cultivées ou des espèces sauvages, soit d'intérêt notoire dans l'alimentation des



**Tableau 1** : production, consommation et conservation des plantes cultivées d'importance alimentaire majeure - a) statistiques de la FAO 1999 <http://apps.fao.org/> et b) données CGIAR 1994 <http://www.noc1.cgiar.org/>

type de culture	production <sup>a</sup> (millions de tonnes en 1999)			consommation <sup>a</sup> (millions de tonnes en 1997)			maintenance <sup>b</sup> (en 1994) espèces cultivées (espèces sauvages)	
	monde	Afrique	Asie	monde	Afrique	Asie	nbre d'accessions	centre CGIAR pays
blé	578	15	257	419	34	252	71 171 7 835	CIMMYT Mexique ICARDA Syrie
maïs	600	38	169	102	30	41	13 070	CIMMYT Mexique
mil	29	13	15	21	9	11	21 392	ICRISAT Inde †
riz	587	18	533	341	13	310	80 646 (2 777) 12 315 (164) 17 440 (880)	IRRI Phillippines IITA Nigeria WARDA Côte d'Ivoire
sorgho	66	20	18	25	14	11	36 719	ICRISAT Inde†
<b>total céréales</b>	<b>2 061</b>	<b>111</b>	<b>1 021</b>	<b>930</b>	<b>105</b>	<b>630</b>		
manioc	165	91	46	93	58	24	2 330 (259) 5 985 (353)	IITA Nigeria CIAT Colombie
pomme de terre	120	9	92	168	7	60	6 257 (1 567)	CIP Pérou
<b>total racines&amp;tubercules</b>	<b>627</b>	<b>154</b>	<b>249</b>	<b>359</b>	<b>92</b>	<b>149</b>		

† : données obtenues pour 1999

populations humaines, soit d'intérêt industriel et économique important. Néanmoins, certaines plantes (et animaux) sont menacés d'extinction suite à une réduction drastique de l'habitat, du nombre et de la taille des populations qui les composent. Dans ces conditions, les décisions de conservation d'espèces ou de taxa menacés de disparition sont prises en fonction des potentialités de ces formes en péril, que ce soit dans le domaine de la pharmacologie ou de l'amélioration des espèces cultivées, ou encore selon d'autres intérêts anthropocentriques comme pour certains mammifères, oiseaux, invertébrés, orchidées, autres plantes à fleur, graminées par exemple.

#### *a) espèces d'importance alimentaire*

De nombreuses collections plus ou moins importantes en taille ont été rassemblées à travers le monde. La FAO (Organisation des Nations Unies pour l'Alimentation et l'Agriculture) a publié un rapport sur l'état des ressources phytogénétiques dans le monde à l'occasion de la Conférence Internationale de Leipzig en Juin 1996.

A titre d'exemple, le rapport de la FAO en 1997 faisait état d'une consommation mondiale de blé de 419 millions de tonnes soit 45% de l'ensemble des céréales consommées (FAO, 1999) (Tableau 1). Par voie de conséquence, depuis 1994, près de 79 000 accessions de blé tendre sont maintenues dans les centres du CGIAR (Groupe Consultatif pour la Recherche Agricole Internationale) (CGIAR, 1999). A l'heure actuelle, le riz est la première céréale d'importance en Asie, loin devant le blé (respectivement 533 et 257 millions de tonnes) (Tableau 1). En 1997, le riz représentait environ la moitié de la consommation céréalière des pays asiatiques. La collection de riz maintenue à l'IRRI (Institut International pour la Recherche sur le Riz) aux Philippines comptait en 1994 plus de 80 000 accessions. En Afrique, parmi les plantes à racines et tubercules, le manioc tient une place essentielle (63%) dans l'alimentation humaine (Tableau 1). En 1994, 8 315 accessions de manioc étaient maintenues dans les banques des centres du CGIAR, et 4 132 accessions étaient disponibles en 1992 au CENARGEN (Centre National de Recherche pour les Ressources Génétiques et les Biotechnologies, Brésil, Cordeiro et al., 1995).

Particulièrement bien adaptées aux zones d'extrême sécheresse, certaines plantes sont essentiellement cultivées dans les zones tropicales semi-arides. Leur contribution dans l'alimentation humaine des régions en voie de développement les rend prioritaires pour les programmes de conservation des ressources génétiques. Tel est le cas pour les cultures de sorgho et de mil, productions céréalières vivrières principales pour les habitants des zones

arides en Afrique et en Inde. La collecte et la conservation des ressources génétiques de ces deux céréales ont été entreprises en 1970 par la FAO et l'IBPGR (Groupe International pour les Ressources Génétiques des Plantes) devenu l'IPGRI (Institut International pour les Ressources Génétiques des Plantes), après une première tentative effectuée 10 ans auparavant par la fondation Rockefeller (Bezançon et al., 1999). Les ressources génétiques de ces deux céréales sont rassemblées dans plusieurs collections nationales et internationales. Les deux collections les plus importantes du point de vue de leur taille et de leur représentativité géographique sont celles de l'ICRISAT (Institut International de Recherche pour les Cultures des régions Tropicales Semi-Arides) en Inde (Tableau 1) et du NPGS (Système National pour le Germplasm des Plantes) aux USA. Nous reviendrons ultérieurement sur les ressources génétiques du sorgho.

*b) espèces cultivées à intérêt économique pour l'agriculture des pays tropicaux*

Utilisées comme fourrage, pour l'industrie textile, la pharmacologie ou encore les productions d'alcool ou de carburants, certaines plantes cultivées ou sauvages présentent des intérêts économiques majeurs en raison de leurs produits primaires ou de leurs dérivés secondaires. Ainsi, l'huile de palme est un produit industriel de haute importance économique. Sa production a été multipliée par 10 en 35 ans, passant en 1995 à 15,6 millions de tonnes (Jacquemard et al., 1997). L'Asie du sud-est joue un rôle majeur dans cette production et notamment la Malaisie, premier producteur avec 26% de la surface totale cultivée et 45% de la production mondiale (FAO, 1999). Le palmier à huile (*Elaeis guineensis*) est aussi la culture qui a connu la plus rapide expansion dans les tropiques. Malgré cet essor important, cette espèce a conservé une base génétique très étroite (Arasu et Rajanaidu, 1975). En 1983, sa conservation a donc été déclarée prioritaire par le conseil de l'IBPGR. Des collections ont été développées notamment au PORIM (Institut de Recherche sur le Palmier à Huile en Malaisie) avec 60 000 accessions (Barcelos, 1998). D'autres collections moins importantes sont également disponibles dans des centres de recherches en Indonésie, Côte d'Ivoire, Nigéria, et au Brésil.

La production mondiale de cacao, fournie essentiellement par l'Amérique centrale et l'Amérique du Sud jusqu'au début du siècle, est aujourd'hui assurée par l'Afrique et notamment la Côte d'Ivoire et le Ghana. Les collectes de cacaoyers ont été organisées depuis les années 30. Les collections les plus importantes comptent entre 700 et 2 500 géotypes et sont conservées à la Trinidad, au Brésil et au Costa Rica (Lanaud et al., 1999).



Le café marchand est une source de devises importantes pour les pays d'Amérique centrale et du Sud, et pour l'Afrique inter-tropicale. D'importantes collections des deux espèces cultivées majeures mais aussi d'autres espèces ont été constituées à Madagascar, en Ethiopie, en Côte d'Ivoire et au CATIE (Centre Agronomique Tropical pour la Recherche et l'Enseignement) au Costa Rica. Dans ce dernier, sont recensés plus de 1 500 accessions de *Coffea arabica* (CATIE, 1999).

Issues du pavot (*Papaver somniferum*), la morphine et la codéine extraites des latex d'opium sont connues pour être des antalgiques très puissants. Ces derniers sont des composants essentiels en pharmacologie et sont largement utilisés dans la médication mondiale. Cette plante a été largement exploitée et depuis les années 70, une collection de 2 000 accessions est maintenue dans la banque de germplasm de l'AARI (Institut Egéen de Recherche Agricole) à Izmir en Turquie. Plus récemment de petites collections ont été établies dans le Nord de l'Inde.

### c) les espèces sauvages

A l'inverse des espèces cultivées, les espèces sauvages n'ont pas été directement soumises aux pressions de sélection humaine. Les pressions de sélection environnementale naturelle ont joué un rôle important dans leur évolution et ont conduit à l'adaptation de ces espèces aux conditions éco-géographiques ou au développement de résistances spécifiques. Les espèces sauvages renferment donc de grandes potentialités et nombre de gènes d'intérêt ont déjà été identifiés, dont certains ont été transférés à des espèces cultivées. Ainsi, le CIP (Centre International sur la Pomme de terre) au Pérou, maintient 1 567 accessions de *Solanum* sauvages (Tableau 1). Dans cette collection, des gènes d'intérêt ont été identifiés pour une large gamme de résistances à des pathogènes, aux insectes nuisibles, pour la tolérance au gel, etc. Ces caractéristiques font qu'aujourd'hui, de nombreuses variétés cultivées intègrent dans leur matériel génétique une portion de génome d'espèce sauvage (Ford-Lloyd, 1986). D'autres collections d'espèces sauvages de pomme de terre ont également été développées en Ecosse, aux USA et en Allemagne (IPGRI, 1995).

L'IRRI aux Philippines, recense 2 777 accessions de riz sauvage (représentant 20 espèces), de même 164 accessions (10 espèces) et 880 accessions (2 espèces) sont conservées respectivement à l'IITA (Institut International en Agriculture Tropicale) au Nigéria, et au WARDA (Association pour le Développement de la Riziculture en Afrique de l'Ouest) en Côte d'Ivoire. C'est à partir de telles collections que des gènes identifiés chez une espèce



sauvage indienne ont permis de protéger les riz cultivés asiatiques de quatre maladies majeures.

De nombreuses collections de tomates ont été développées de par le monde (Chine, Cuba, Allemagne, Hongrie, Hollande, Pérou, USA et Russie) (IPGRI, 1995). Elles intègrent de nombreuses espèces sauvages essentielles pour l'amélioration des variétés cultivées : résistance aux stress environnementaux (sols salés, sécheresse), aux attaques d'insectes, à des pathogènes; amélioration des rendements, tenue du fruit (gène pour prévenir l'abscission du fruit), qualité du fruit, teneur en sucre, capacité du fruit à résister aux manipulations et au stockage.

Aujourd'hui, on peut considérer que la collecte des plantes cultivées alimentaires et d'intérêt économique est bien avancée. Par contre celle des espèces sauvages reste encore dans nombre de cas à promouvoir. Quelles espèces doivent être collectées ? Des priorités ont été définies ; elles concernent les espèces à priori les plus facilement utilisables et celles menacées d'extinction (Chapman, 1989).

### III - Conservation des ressources génétiques

Depuis les années 70, un réseau international pour la conservation des ressources phytogénétiques a été développé et est administré conjointement par le CGIAR, l'IPGRI et la FAO. Suite à la conférence de la Terre tenue à Rio de Janeiro en 1992, la convention sur la diversité biologique (CBD) a été signée par 150 gouvernements. Trois objectifs principaux ont été fixés : 1) promouvoir des stratégies efficaces pour la conservation des ressources génétiques, 2) utiliser durablement la biodiversité et 3) partager équitablement les bénéfices de l'utilisation des ressources génétiques.

La conservation des ressources génétiques nécessite la préservation de la diversité génétique en considérant la diversité des espèces et des écosystèmes associés. Par conséquent, les programmes de conservation des ressources génétiques doivent prendre en compte la survie à long terme des espèces et le maintien des processus écologiques et évolutifs (Hamrick et Godt, 1995).

La conservation des espèces en danger d'extinction, des écotypes qui présentent des propriétés pharmacologiques, ou des formes esthétiquement remarquables, est

**Tableau 2** : banques de gènes et nombre d'accessions maintenues dans les collections nationales et régionales *ex situ*, par région (Sources : *Rapports nationaux et base de données SIAM*, IPGRI, 1996).

région	entrées		banque de gènes	
	nombre	(%)	nombre	(%)
Afrique	353 523	(6)	124	(9)
Amérique latine et Caraïbe	642 405	(11)	227	(17)
Amérique du Nord	762 061	(13)	101	(8)
Asie	1 533 979	(27)	293	(22)
Europe	1 934 574	(34)	496	(37)
Proche-Orient	327 963	(6)	67	(5)
<b>total</b>	<b>5 554 505</b>		<b>1308</b>	
<b>total GCIAR</b>	<b>593 191</b>		<b>12</b>	

essentiellement menée dans des jardins botaniques, des arboretum ou des conservatoires horticoles.

Dans ce qui suit, nous nous intéresserons exclusivement aux collections portant sur les ressources génétiques des plantes cultivées. Deux stratégies complémentaires sont développées pour assurer la conservation des ressources phytogénétiques : la conservation *ex-situ* et la conservation *in-situ*.

### 1) conservation *ex-situ*

La conservation *ex-situ* à court, moyen et long terme, est gérée dans des centres de recherche pour maintenir des formes cultivées et/ou sauvages, plus ou moins éloignées de leurs sites de culture originels. Les collections *ex-situ* ont été développées pour la majorité des plantes cultivées majeures. Elles incluent les variétés traditionnelles, des espèces sauvages apparentées, des lignées améliorées et des stocks génétiques. Depuis les années 70, de nombreuses banques de gènes gérées par des centres de recherche nationaux et internationaux ont été constituées (Tableau 2). La conservation *ex-situ* se fait selon plusieurs modalités : le plus souvent, banque de graines (court, moyen et long terme), parfois pollen congelé ou encore, pour les plantes à reproduction végétative, microbouturage *in-vitro* et/ou cryo-conservation de tissus somatiques (apex ou embryon). Dans un certain nombre de cas, on dispose aussi depuis quelques années de banque d'ADN. Certaines espèces ne se prêtent pas à la conservation par graines et nécessitent des collections *ex-situ* vivantes. C'est le cas des collections d'arbres ou encore de plantes à reproduction végétative telles que le bananier, le manioc ou encore les ignames.

Bien que les collections *ex-situ* soient appelées "collections statiques" (Bennett, 1968 cité par Astley, 1992), des changements délétères et des mutations au cours du stockage peuvent se produire (Pernès, 1984). Des travaux portant sur des collections de graines de laitue, orge et pois, ont permis de montrer une relation étroite entre la viabilité et l'accumulation de dommages chromosomiques (Pernès, 1984 ; Astley, 1992). De même, tous les génotypes des différentes espèces de caféiers ne répondent pas de manière équivalente à la conservation par microbouturage *in vitro*. En quelques années, des populations, voire des espèces, peuvent être totalement perdues (Dussert et al., 1997).

Les collections *ex-situ* offrent donc l'avantage de pouvoir préserver dans un lieu - le plus sécurisé possible - le maximum de ressources génétiques. Cependant, ce mode de conservation n'est pas sans risque - mauvaise réponse aux conditions de conservation,



modifications du matériel héréditaire, absence de co-évolution plante/environnement (au sens le plus large) - et ne peut donc être le seul envisagé.

## 2) conservation *in-situ*

La conservation *in-situ* permet de maintenir des populations dans leurs conditions "naturelles" d'évolution, avec un système agro-écologique traditionnel préservé, et un environnement socioculturel associé. Ce mode de conservation favorisant le maintien des espèces sauvages, des variétés forestières et fruitières, des espèces à multiplication végétative, à graines récalcitrantes ou dont les conditions de stockage au froid sont non définies, dans leur contexte évolutif naturel est aussi qualifié de "conservation dynamique". L'avantage de telles collections réside dans la possibilité d'accéder à des variétés en co-évolution continue avec les agents pathogènes (bactéries, virus ou champignons), et en adaptation permanente avec leur environnement éco-géographique. Cependant, ne permettant un accès direct aux gènes utiles, l'inconvénient majeur des collections *in-situ* reste leur intérêt parfois limité pour les programmes de sélection (Brush, 1995). Une telle limitation est souvent responsable du faible intérêt porté par les centres de recherches à un tel mode de conservation.

## 3) conservation "on-farm"

Parmi les stratégies de conservation *in-situ*, la conservation "on-farm" ou "maintien en culture" (Louette, 1994) fait intervenir une approche participative des paysans. Dans ces conditions, les variétés cultivées, maintenues dans leur agroécosystème originel, sont gérées par les paysans eux-mêmes perpétuant leurs propres méthodes et critères de sélection.

La conservation on-farm constitue une situation de compromis dans laquelle peuvent être conciliés les objectifs de conservateurs, d'écologistes et du développement économique et social (Jarvis, 1999). Cependant, la conservation on-farm n'est pas toujours réalisable et ce, en grande partie, pour des raisons socio-économiques. En effet, il est difficilement concevable de penser que les paysans sont plus sensibles à la réduction de la diversité génétique qu'à la réduction de leur pouvoir d'achat. En tout état de cause, pour le cultivateur, la conservation on-farm doit être source de revenus financiers sans quoi elle ne peut pas être mise en place. Des scientifiques mais surtout des politiques doivent participer à la réflexion pour définir les moyens à mettre en oeuvre pour obtenir ce profit.



Quelques espèces sont concernées par de telles collections. Ainsi, la conservation on-farm est pratiquée dans des fermes péruviennes pour le maintien de variétés traditionnelles de pommes de terre, également dans le sud du Mexique pour des variétés de maïs, dans l'ouest de la Turquie et concerne le blé (Brush, 1995) ou encore au Népal pour le riz.

Finalement, la gestion des collections, *ex-situ*, *in-situ* ou encore on-farm, n'est pas sans poser des problèmes. La conservation même à long terme nécessite la régénération. Celle-ci a lieu dans des conditions contrôlées bien définies (Engels et Ramanatha Rao, 1995) et des difficultés majeures sont rencontrées chez certaines espèces sauvages (Brown et al., 1997). Comment diminuer les coûts de maintenance et de régénération dans les banques de gènes *ex-situ* ? Comment organiser au mieux la gestion des collections *in-situ* ? Comment développer - sans contradiction - la conservation on-farm ? Peut-on définir des priorités de conservation ? Faut-il concentrer les efforts sur une portion de la collection ? Comment procéder alors au choix des accessions prioritaires ? Pour apporter des éléments de réponse à ces questions, la connaissance du matériel végétal en conservation est indispensable. Celle-ci est facilitée par l'évaluation des ressources génétiques.

#### IV - Evaluation des ressources génétiques

Après la collecte du matériel, et quand les efforts nécessaires ont été mis en oeuvre pour assurer le maintien et la disponibilité des accessions conservées, la description de la collection est un des pré-requis à l'utilisation optimale des ressources génétiques. S'en suit donc une étape d'évaluation - plus ou moins importante et complète - des collections.

Dans les collections de grande taille, il n'est pas concevable de procéder à la caractérisation morpho-agronomique simultanée de toutes les accessions (i.e. dans les mêmes conditions éco-environnementales). Généralement, le matériel végétal est caractérisé sur plusieurs années, le plus souvent à l'aide des descripteurs morphologiques recommandés par l'IPGRI.

Des dispositifs expérimentaux précis sont nécessaires à 1) l'évaluation de certains caractères 2) la prise en compte de l'effet environnemental. S'agissant d'études extensives de toute la collection simultanément, la mise en place de tels dispositifs constitue une contrainte très forte. Ainsi, seul un nombre restreint d'accessions donne lieu à des évaluations agronomiques et/ou de résistance à des stress biotiques et abiotiques. Une fois de plus, se pose la question de savoir comment choisir les accessions qui seront prioritairement évaluées.

Lorsque des évaluations sont régulièrement réalisées, les informations collectées sont mises à disposition dans des bases de données informatisées telles que GRIN (Réseau d'Information sur les Ressources Génétiques) aux USA, SINGER (Réseau d'Information pour les Ressources Génétiques) en Italie, le GRID (base de Données pour l'Information des Ressources Globales) du programme environnement des Nations Unies, IGMS (Système de Gestion de la Banque de Gènes) de l'ICRISAT, SISTER (Système Informatisé pour le Stockage, le Traitement et l'Evaluation des Résultats) du CIRAD,... La collection ainsi documentée devient théoriquement plus accessible pour les programmes de sélection en permettant de cibler dans l'ensemble du germplasma, les accessions présentant un caractère particulier. Cette dernière étape suppose au préalable l'analyse des données de la base.

## V - Core collection

Nous avons vu précédemment que les collections *ex-situ*, de par leur taille, posent des problèmes pour leur gestion et leur évaluation. Par voie de conséquence, l'utilisation des ressources génétiques paraît être loin de l'optimum.

Ainsi, pour la maintenance comme pour l'évaluation du germplasma, la distinction entre "collection de réserve" et "collection active" paraît être une solution pour assurer une meilleure exploitation et une plus grande disponibilité des ressources génétiques. La collection de réserve ou encore collection de base joue le rôle de conservation à long terme, et sert à fournir des semences à la collection active.

La collection active correspond à un ensemble de taille réduite par rapport à la collection de base et représentatif de celle-ci. Elle est utilisée pour permettre l'évaluation d'un grand nombre de caractères et par la suite pour faciliter l'évaluation raisonnée de la collection de base. Comment constituer un tel ensemble ? C'est dans cette problématique que le concept de "core collection" ou "collection noyau" est apparu (Frankel, 1984, cité dans Brown, 1989a).

### 1 - Définition

Une "core collection" se définit comme un ensemble d'accessions, le moins redondant possible, et représentant au mieux la diversité génétique de la collection de base. La représentativité d'une core collection peut être estimée par :

- l'importance de la diversité génétique globale contenue dans la core collection par rapport à celle de la collection de base;

- la ressemblance entre la distribution de la diversité de la core collection et celle de la collection de base (Galwey, 1995).

## 2 - Stratégies de constitution de core collections

A partir d'une hypothèse de neutralité des gènes, et en supposant l'interfertilité de toutes les accessions contenues dans la collection de base, Brown, (1989a) démontre qu'un prélèvement aléatoire de 10% sur l'ensemble initial devrait permettre de retenir jusqu'à 70% de la diversité génétique globale. Depuis, l'intérêt grandissant porté à la constitution de core collections a conduit à la proposition de nombreuses stratégies d'échantillonnage. Dans la plupart des cas, c'est la disponibilité des informations sur le matériel étudié qui détermine le mode de prélèvement.

### 2 - 1. Stratégies aléatoires

Certaines core collections sont constituées par prélèvement aléatoire directement à partir de la collection de base ("stratégie R" ; random). Pour d'autres, l'échantillonnage fait suite à une étape préliminaire d'analyse de la collection de base. Celle-ci a pour but de mettre en évidence l'existence éventuelle d'une structuration forte et donc d'en tenir compte lors du prélèvement par un échantillonnage stratifié.

#### 2 - 1. 1. structuration de la collection de base

Généralement, peu de données d'évaluation sont disponibles pour les collections de base. Dans ce cas, les données passeport (critères éco-géographiques) peuvent être utilisées pour hiérarchiser la collection de base. De telles stratifications ont été utilisées pour la constitution des core collections de soja sauvage et orge (Brown, 1989b), blé dur (Spagnoletti Zeuli et Qualset, 1993), luzerne pérenne des USA (Basigalup et al., 1995) et orge collectés en Chine (van Hintum et al., 1995b). La stratification peut être aussi agro-écologique (basée sur les enregistrements des rendements moyens par province de 47 années) comme dans la collection d'orge espagnole (Igartua et al., 1998), spatiale dans le cas des collections de luzernes annuelles de Méditerranée (Balfourier et al., 1998) ou encore biologique (taxonomie, caractéristiques reproductives) dans la collection de luzerne annuelle australienne (Skinner et al., 1999).

Avec une documentation plus riche des collections, davantage de critères peuvent être considérés pour mettre en évidence une organisation de la diversité. Combinant des critères



géographiques et morpho-agromomiques, des structurations ont été révélées dans les collections de gombo (Hamon et van Sloten, 1989), d'arachide (Holbrook et al., 1993), de luzerne pérenne (Charmet et al., 1994), de luzerne annuelle des USA (Diwan et al., 1994 et 1995), de manioc du Brésil (Cordeiro et al., 1995), de maïs de la race Tuxpeño (Crossa et al., 1995), de sorgho (Prasada Rao et Ramanatha Rao, 1995), de haricot (Thome et al., 1995), d'espèces fourragères (Balfourier et al., 1998), de quinoa (Ortiz et al., 1998) et dans la collection de patate douce (Huamán et al., 1999).

Rares sont les collections actuellement évaluées à l'aide de marqueurs biochimiques ou moléculaires. Cependant, à partir d'un sous-ensemble caractérisé à l'aide des marqueurs RAPDs (polymorphisme d'ADN amplifiés au hasard) une core collection de choux devrait être développée par échantillonnage dans des groupes prédéfinis à l'aide de marqueurs RAPDs (Divaret et al., 1999).

## 2 - 1. 2. échantillonnage

A partir de la collection stratifiée, et lorsqu'aucun autre critère n'est pris en compte, des échantillonnages aléatoires peuvent être effectués par prélèvement d'un nombre d'accessions

- constant au sein de chacun des groupes ("stratégie C") ;
- proportionnel à la taille des groupes ("stratégie P") ;
- proportionnel au logarithme de la taille des groupes ("stratégie L"), (Brown, 1989b).

Des modèles théoriques, intégrant une caractérisation biochimique ou moléculaire de la collection, ont conduit à d'autres stratégies de prélèvement après stratification. Sous l'hypothèse d'un modèle d'allèles neutres, et lorsque la diversité de la collection se structure en groupes génétiquement isolés, un échantillonnage peut être réalisé en prenant en compte le paramètre de diversité génétique ("stratégie H", Schoen et Brown, 1993). A partir des estimations de la diversité intra-groupe (définie selon la formule de Nei  $H = 1 - \sum(p_i)^2$ ), cette stratégie permet de définir le nombre d'accessions devant être prélevées au sein de chaque groupe.

Une autre stratégie aléatoire après stratification ("stratégie G"), est proposée par Yonezawa et al. (1995). Elle permet d'effectuer un prélèvement proportionnel à la diversité génétique intra-groupe. Ici, la diversité est mesurée par le nombre de classes de compositions alléliques trouvées dans les différents groupes structurant la collection.



Des simulations de constitution de core collections ont montré la supériorité de ces deux stratégies (H et G) pour maximiser la richesse allélique par rapport aux stratégies qui ne tiennent compte d'aucune information quant à la répartition de la diversité intra-groupe (Schoen et Brown, 1995 ; Bataillon 1994 ; Yonezawa et al., 1995).

## 2 - 2. Stratégies non-aléatoires

A partir de collections bien documentées, différentes stratégies d'échantillonnage non aléatoire peuvent servir à la constitution de core collections. Les accessions sont sélectionnées sur des critères phénotypiques, biochimiques ou moléculaires, pour maximiser la représentation de la diversité contenue dans la collection de base.

### 2 - 2. 1. critères phénotypiques

A partir d'analyses multivariées portant sur des caractères morpho-agronomiques, des core collections de blé dur (Spagnoletti Zeuli et Qualset, 1993 suivant une stratification à la fois phénotypique et géographique), luzerne des USA (Basigalup et al., 1995 après une classification hiérarchique sur les composantes principales), maïs (Rincon et al., 1997 par une analyse en composantes principales, et Taba et al., 1998 sur index de sélection) et luzerne (Skinner et al., 1999 par une sélection sur les distances Euclidiennes) ont été constituées.

Diwan et al. (1995) ont développé une stratégie appelée "diversité relative" combinant à la fois les données passeport et phénotypiques pour la constitution de la core collection de luzerne annuelle des USA. Dans une première étape, des groupes sont déterminés à partir des données phénotypiques et/ou passeport. Le nombre d'accessions à prélever est déterminé en fonction de la diversité intra-espèce elle-même définie par le nombre de groupes obtenus dans chaque espèce. Le niveau des groupes à considérer est fonction de la taille de la core collection fixée, sachant qu'une accession par groupe sera retenue. L'étape finale consiste en un prélèvement plus ou moins empirique, visant à maximiser la représentation géographique.

La stratégie PCS (Principal Component Scoring) a été développée pour maximiser la représentation de la diversité phénotypique au sein de la core collection (Noirot et al., 1996). Sous l'hypothèse d'absence de compartimentation génétique et considérant que deux phénotypes extrêmes peuvent redonner par recombinaison génétique la gamme des phénotypes intermédiaires, cette stratégie permet de sélectionner les accessions qui contribuent le mieux à la variabilité phénotypique. Cette stratégie peut être appliquée à des données quantitatives (Hamon et al., 1995 chez les caféiers ; Mahajan et al., 1996 pour la

collection de gombo sud asiatiques ; Bisht et al., 1998 pour la collection indienne de haricot) ou qualitatives (Deu et al., 1999 pour une collection de sorgho). L'application de cette stratégie a également été étendue en utilisant des composantes "spatiales" issues de géostatistiques multivariées (Balfourier et al., 1998 pour les collections de luzerne).

### *2 - 2. 2. critères biochimiques ou moléculaires*

Les critères moléculaires ou biochimiques peuvent également être pris en compte par les stratégies de constitution de core collection. Les données moléculaires sont des données qualitatives pouvant être utilisées dans une stratégie PCS (Barcelos, 1998 pour le palmier à huile de la collection brésilienne; Hamon et al., 1998 pour les collections de caféiers, riz, hévéa et sorgho). A partir de 155 accessions de cacaoyers évaluées à l'aide de marqueurs RFLPs et RAPDs, Lerceteau et al. (1997) ont suggéré la constitution d'une core collection sur la base des indices de diversité moléculaire. L'échantillonnage proposé est basé sur la stratégie développée par Diwan et al. (1995). Sur une collection structurée géographiquement, la "stratégie M" prélève au sein de groupes, en prenant en compte la diversité moléculaire. Les accessions les plus polymorphes en terme de richesse allélique et très bien différenciées les unes par rapport aux autres, sont retenues (Schoen et Brown, 1993).

Des simulations de constitution de core collections ont montré la très nette supériorité de la stratégie M, par rapport aux autres modes de prélèvement H, P, L, C et R pour retenir la diversité génétique (Schoen et Brown, 1993; Bataillon, 1994; Bataillon et al., 1996, Schoen et Brown, 1995).

### *2 - 2. 3. autres critères*

D'autres stratégies d'échantillonnage sélectif peuvent être proposées, en basant la sélection d'accessions sur un ensemble de critères différent. A partir des coefficients de parenté entre cultivars modernes d'orge, des accessions ont été sélectionnées de façon à maximiser le nombre de lignées parentales distinctes retenues (van Hintum et Haalman, 1994). D'après des caractéristiques éco-géographiques ou agro-morphologiques, des informations relatives aux modes de culture, aux environnements socio-culturels etc., des sélections très subjectives peuvent être effectuées. Les critères de sélection pris en compte dans ce type de prélèvement peuvent être variables et modulés selon les objectifs assignés à la core collection. Cette stratégie implique une connaissance de la plante et de son

environnement (dans son sens le plus large) très importante que seule l'expérience de terrain durant de nombreuses années peut fournir.

Face à cette diversité de stratégies proposées, des questions se posent quant au choix de la méthode à utiliser pour constituer une core collection. Il s'agit alors : - de définir une stratégie en fonction de la nature des données accessibles. Par exemple la "stratégie H" ne sera pas applicable à une collection évaluée pour des caractéristiques morphologiques ; - d'étudier les relations entre diversité moléculaire et diversité phénotypique, cette dernière répondant directement aux exigences actuelles des utilisateurs ; - de connaître les conséquences des différentes stratégies sur les caractéristiques des core collections obtenues.

### 3 - Utilisation des core collections

Une core collection n'est pas une entité en soi. Elle doit plutôt être considérée comme un point d'entrée dans la collection de base (Hodgkin et al., 1995; Vaughan et Jackson, 1995). Après sa constitution, la représentativité de la core collection vis-à-vis de la collection de base doit être testée. Que ce soit par l'utilisation des indices de diversité tels que l'indice de diversité phénotypique global de Shannon-Weaver ( $-\sum p_i \log_e p_i$ ), ou par des comparaisons de distribution des caractères qualitatifs, ou des comparaisons de variances et moyennes pour les caractères quantitatifs, ce sont surtout les données morpho-agronomiques, généralement les premières disponibles, qui sont prises en compte. Ainsi, une bonne représentation de la variabilité phénotypique de la collection de base a été vérifiée pour certaines core collections (Holbrook et al., 1993; Spagnoletti Zeuli et Qualset, 1993 ; Casler, 1995 ; Charmet et Balfourier, 1995 ; Cordeiro et al., 1995 ; Galwey, 1995 ; Bisht et al., 1998 ; Mahajan et al., 1996 ; Balfourier et al., 1998 ; Ortiz et al., 1998 ; Huamán et al., 1999). Dans d'autres cas, les core collections ont montré une divergence très nette par rapport à la collection de base. Cette différence se traduit essentiellement par une augmentation de la variance aux caractères quantitatifs, due à la stratégie d'échantillonnage (Spagnoletti Zeuli et Qualset, 1993 pour d'autres core collections ; Diwan et al., 1994 et 1995; Basigalup et al., 1995 ; Jung et al., 1997 ; Skinner et al., 1999).

Lorsque les données moléculaires ne sont pas complètement disponibles, il n'est pas possible de rapporter directement la diversité moléculaire de la core collection à celle de la collection de base. Par contre, il est possible de comparer les diversités moléculaires de deux



**Tableau 3** : exemples de core collections évaluées pour des caractères d'intérêt.

espèce	core collection	évaluation	référence
arachides	Holbrook et al., 1993	résistance à la leafspot	Holbrook et Anderson, 1995
papate douce péruvienne	Huamán et al., 1999	six traits d'intérêt morpho-agronomique (résistances)	Huamán et al., 1999
alfalfa pérenne des USA	Basigalup et al., 1995	résistance à l'acidité des sols qualités fourragères	Bouton, 1996 Jung et al., 1997
trèfle blanc	Perderson et al., 1996	cyanogénèse	Perderson et al., 1996
quinoa péruvienne	Ruiz-Tapia et al., 1998†	tolérance au sel	Ortiz et al., 1999
choux	Boukema et al., 1997‡	résistance aux aphides des crucifères	Ellis et al., 1998

† cité dans Ortiz et al., 1999

‡ cité dans Ellis et al., 1998



échantillons pris aléatoirement au sein de la core collection et de la collection de base. Ceci a été réalisé chez le haricot. Un prélèvement aléatoire dans la core collection constituée par Thome et al. (1995), et un second échantillon aléatoire issu de la collection initiale ont été caractérisés et évalués à l'aide de marqueurs RAPDs (Skroch et al., 1998). Aucune différence significative n'a été notée concernant la diversité moléculaire globale (fréquences alléliques, diversité génétique moyenne, distances génétiques). Un tel constat a permis aux auteurs de vérifier que par la procédure mise en oeuvre pour la constitution de cette core collection, la diversité n'était pas plus perdue que par un échantillonnage au hasard.

Par la suite, la core collection peut être évaluée pour des caractères spécifiques et ainsi permettre un meilleur accès à la collection de base (Tableau 3). La core collection d'arachides, constituée par Holbrook et al. (1993) à partir d'une collection de base stratifiée en fonction des pays d'origine (et des données morphologiques pour les cas où de telles informations étaient disponibles), a été testée vis-à-vis de la résistance au virus du leafspot. La collection de base ayant été entièrement évaluée pour cette résistance, la core collection est validée rétrospectivement en déterminant son efficacité à permettre l'identification des sources de résistance. Deux étapes ont été réalisées : 1) l'étude de l'intégralité de la core collection de façon à repérer les groupes (issus de la hiérarchisation préalable) susceptibles de présenter un bon taux de résistance, 2) un retour à la collection de base et l'étude extensive des groupes "prometteurs". Les auteurs ont ainsi identifié 54% des génotypes de la collection de base présentant une résistance, à partir de l'étude ciblée de 27% des accessions de la collection de base (Holbrook et Anderson, 1995).

De plus, une documentation exhaustive de la core collection et l'étude des corrélations entre les caractères permettent d'identifier les informations redondantes, et ainsi de simplifier les évaluations de la collection de base (Skinner et al., 1999).

Finalement, les core collections doivent être considérées comme des ensembles dynamiques, susceptibles d'évoluer en fonction de récentes acquisitions d'accessions, de données, ou encore des objectifs assignés par les utilisateurs. Mackay (1995) envisage la constitution de plusieurs core collections pour répondre aux différentes attentes des utilisateurs. Une ou des core collections ? La question réside davantage dans la manière de constituer cette ou ces core collections et dans l'utilité/l'utilisation réelle de cette/ces core collection(s) face à des ressources génétiques dont on ne connaît pas toutes les potentialités à

**Tableau 4** : surface en culture et production pour des plantes cultivées d'importance alimentaire majeure dans le monde, en Afrique, en Asie et en Inde (statistiques de la FAO 1999 <http://apps.fao.org/>)

type de culture	superficie (milliers d'ha.)				production (milliers de tonnes)			
	monde	Afrique	Asie (sauf Inde)	Inde	monde	Afrique	Asie (sauf Inde)	Inde
blé	214 199	8 578	70 029	27 356	578 337	15 495	186 322	71 007
riz	153 120	7 990	93 556	43 000	586 787	17 633	405 376	127 600
maïs	139 190	24 779	37 627	6 300	599 708	38 772	158 891	10 500
orge	58 596	4 369	12 436	783	133 420	4 178	21 050	1 477
sorgho	44 913	22 981	2 637	11 200	65 810	19 702	6 853	11 000
mil	37 242	29 728	2 947	13 100	29 055	12 884	4 734	10 300
<b>total</b>	<b>681 374</b>	<b>90 950</b>	<b>222 449</b>	<b>101 739</b>	<b>1 993 117</b>	<b>108 664</b>	<b>783 226</b>	<b>231 884</b>

un instant donné. A ce point du débat, il s'agit aussi de ne pas confondre l'utilité d'une core collection - pour l'évaluation exhaustive d'un sous-ensemble représentatif de la diversité de la collection - à celle de la collection de travail constituant une base pour des recherches ciblées sur des caractéristiques particulières.

## **B - LE SORGHO**

### **I – La production en sorgho**

Parmi les céréales, le sorgho est en cinquième position en terme de surface cultivée après le blé, le riz, le maïs et l'orge. En 1999, la surface mondiale cultivée en sorgho était de 45 millions d'hectares, avec une production de 66 millions de tonnes (FAO, 1999) (Tableau 4). Dans les zones tropicales semi-arides, l'intérêt porté aux plantes bien adaptées aux conditions extrêmes de sécheresse oriente la céréaliculture vers la production de sorgho. Environ 90% des surfaces consacrées au sorgho se situent dans les pays du Sud où sont cultivées des variétés locales à faible rendement (1,2 t/ha en moyenne contre 4t/ha pour des lignées améliorées aux USA) et à rendement variable selon les années. Plus de 70% de cette production sont destinés à la consommation humaine. Ces sorghos sont également utilisés en tant que complément fourrager pour le bétail, pour la production d'alcools ou encore comme source d'énergie ou matériau de construction.

Une large gamme de photopériodisme permet une bonne adaptation de la culture des sorghos aux zones tempérées. Par conséquent, dans les programmes d'amélioration, de nombreuses conversions de matériel photosensible en photo-insensible ont été effectuées ces 10 dernières années (Prasada Rao et Gopal Reddy, 1989; Doggett et Prasada Rao, 1995). Dans les pays développés, la culture du sorgho est assurée par des variétés cultivées à haut rendement (de l'ordre de 4 t/ha). Cette céréaliculture représente 23% de la production mondiale des sorghos cultivés. Bien que principalement orientées vers une production fourragère, quelques utilisations du sorgho ont été développées dans des filières industrielles. Dans les années 80, l'Université du Nebraska (USA) a développé une ferme expérimentale pour la production de sorgho à des fins de production d'énergie renouvelable (Collectif, 1996). Au Brésil en 1993, 40% du parc automobile ont été adaptés au fonctionnement à l'éthanol obtenu à partir de la canne à sucre. A l'heure actuelle, des chercheurs tentent d'utiliser la production de sorghos sucrés pour une telle industrie.





## II - Taxonomie

Le sorgho - *Sorghum bicolor* L. Moench - est une poacée (ex-graminée), appartenant à la sous-famille des Panicoïdés et à la tribu des Andropogonées (Figure 1). Suivant une classification basée sur la notion de pool génique définie par Harlan et de Wet (1971), le genre *Sorghum* se divise en cinq sections, parmi lesquelles se trouve la section *sorghum* composée de trois espèces ; deux espèces sauvages et pérennes (*S. halepense*, *S. propinquum*) et une espèce annuelle (*S. bicolor*). Cette dernière renferme les sorghos cultivés (ssp. *bicolor*), les sorghos adventices (ssp. *drumondii*) et les sorghos sauvages (ssp. *arundinaceum*). Toutes deux diploïdes ( $2n = 2x = 20$ ), *S. bicolor* et *S. propinquum* partagent le pool génique primaire. Bien qu'isolées géographiquement, ces deux espèces sont interfertiles. *S. halepense*, constituant le pool génique secondaire, est une espèce tétraploïde ( $2n = 4x = 40$ ) pouvant s'hybrider avec les sorghos cultivés pour donner l'espèce *S. alnum* (Doggett, 1970 cité par Morden et al., 1990).

Le sorgho est une plante préférentiellement autogame (en moyenne, de 5 à 7% d'allogamie), bien que certaines formes puissent présenter des taux d'allofécondation allant jusqu'à 30% (Ollitrault, 1987 ; Chantereau et Kondombo, 1994 ; Pedersen et al., 1998).

Différentes clés taxonomiques ont permis de classer les sorghos cultivés en races, sous-races et écotypes (Snowden, 1936 ; Murty et Govil, 1967 ; de Wet, 1978 ; Prasada-Rao et al., 1989). La classification proposée par Harlan et de Wet (1972) est basée sur les critères morphologiques de structure de l'épillet à maturité et de forme de la panicule. Ces critères sont supposés stables et peu soumis aux effets environnementaux. Cette classification a conduit à définir cinq races de base (*bicolor*, *caudatum*, *durra*, *guinea*, *kafir*) et dix formes intermédiaires. Les cinq races de base des sorghos cultivés sont présentées dans la figure 2 et leur aire privilégiée de distribution dans la figure 3.

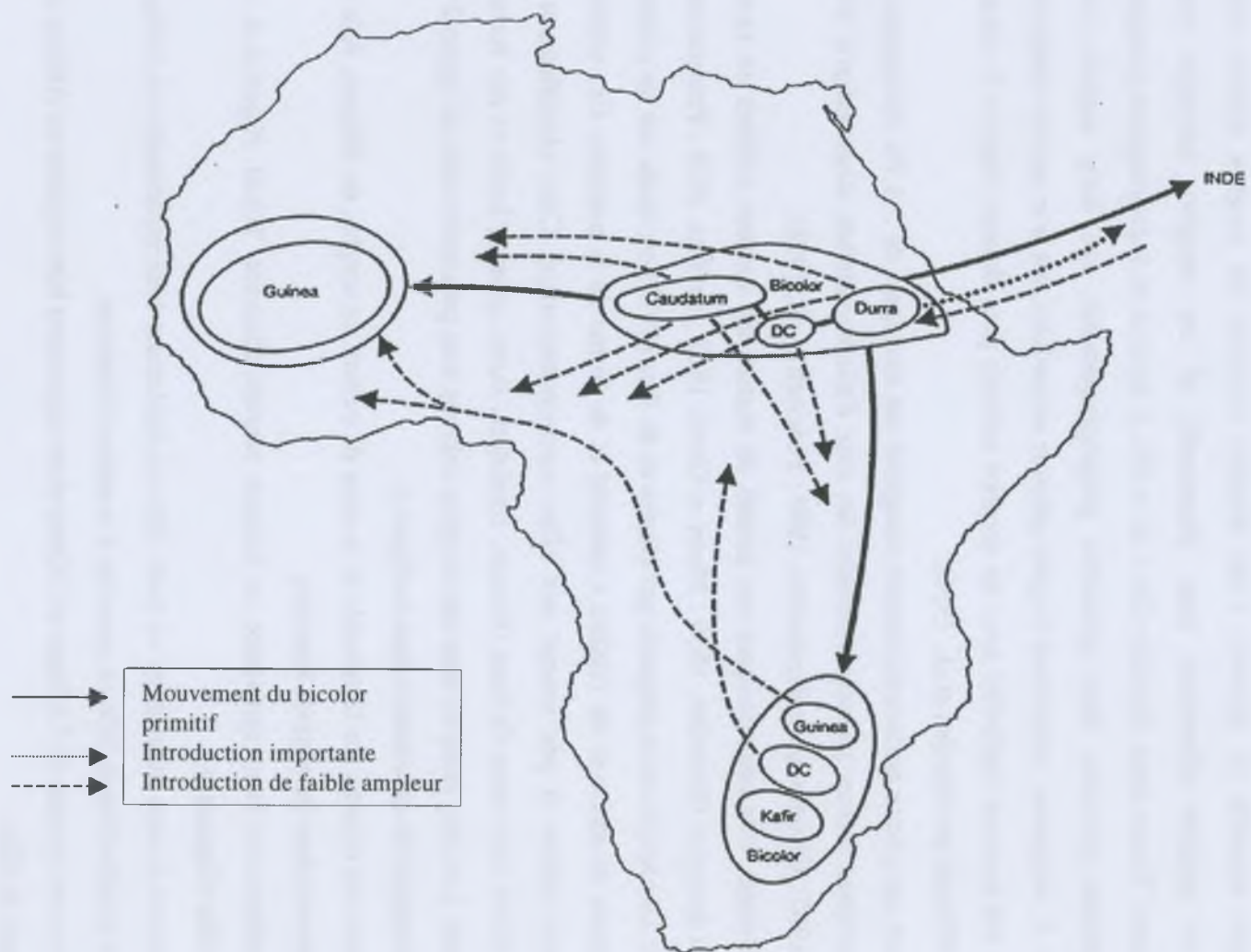
La race *bicolor* est répartie sur l'ensemble de la zone de culture du sorgho, en Afrique, Asie et Chine, et souvent dans les régions humides.

La race *caudatum* est bien représentée en Afrique centrale (Soudan, Tchad, Nigeria) et en Afrique de l'Est (Ouganda).

La race *durra* est la race majoritaire en Inde. Elle est également bien représentée en Afrique de l'Est et au proche-Orient. Elle est associée à la culture islamique.

La race *guinea* est typique de l'Afrique de l'Ouest et est également bien répartie en Afrique du Sud et Afrique de l'Est.





**Figure 4 :** Schéma de domestication des sorghos cultivés (d'après Ollitrault et al., 1989).



La race kafir est essentiellement présente en Afrique du Sud et Afrique de l'Est (au sud de la Tanzanie).

### III - Origine géographique et domestication

Des études morphologiques comparatives et des modèles de distribution géographique des races conduisent à estimer la domestication du sorgho à 3000 ans avant J.C. dans la région du quadrant Nord-Est de l'Afrique, dans une zone de savane en bordure du Sahara (de Wet et Huckabay, 1967 ; Mann et al., 1983). Cependant, des restes de grains carbonisés âgés de 8000 ans, découverts dans des sédiments datant de l'Holocène d'un site archéologique proche de la bordure Egyptio-Soudanaise, laissent supposer une origine du sorgho plus ancienne (Wendorf et al., 1992). Quoi qu'il en soit, tous les auteurs s'accordent à dire que la domestication des sorghos cultivés s'est faite à partir de la sous-espèce de sorghos sauvages *S. bicolor ssp arundinaceum*. Néanmoins, deux hypothèses sur l'origine des races actuelles de sorgho ont été émises :

- une origine polyphylétique basée sur des domestications indépendantes des races cultivées (Snowden, 1936 et de Wet et Huckabay, 1967). Cette hypothèse suggère trois centres de domestication africains, plus ou moins indépendants, qui auraient permis, à partir de trois races sauvages de la sous-espèce *arundinaceum*, le développement distinct des races guinea d'une part, caudatum et kafir d'autre part, et enfin durra ;
- une origine monophylétique avec une primo-domestication des sorghos primitifs, de l'Ethiopie au Niger en bordure sud du Sahara (Harlan et Stemler, 1976). Cette forme primitive serait proche du type actuel bicolor. Puis ces sorghos auraient été introduits en Afrique de l'Ouest, Afrique du Sud et en Inde où ils se seraient différenciés pour donner respectivement les races guinea, kafir et durra. Selon cette hypothèse, la race caudatum se serait différenciée ultérieurement, à proximité de la zone de domestication primaire. Puis, par des événements de migration la distribution des différentes races de sorgho se serait étendue sur l'Afrique et l'Asie.

L'utilisation des outils biochimiques et moléculaires a permis de trancher en faveur d'une origine monophylétique (Ollitrait et al., 1989 ; Aldrich et Doebley, 1992 ; Aldrich et al., 1992 ; Deu et al., 1994 ; 1995). Cependant, si tous les auteurs s'accordent sur l'origine monophylétique des sorghos, certains proposent une origine est-africaine des durra (Ollitrait, 1987) (Figure 4).



**Figure 5 :** distribution mondiale des cultures de sorgho (d'après FAO, 1996).

La présence de sorgho aux Amériques est très récente (1853) et est due à des importations via la France (Doggett, 1988).

#### IV - Les collections de sorgho

La culture du sorgho couvre une vaste aire géographique et se pratique dans tous les continents (Figure 5). Il en résulte une forte diversité des formes en culture associée à une forte diversité du contexte socio-culturel et de l'utilisation des formes cultivées. L'objectif des collectes de variétés traditionnelles est de rassembler le maximum de formes cultivées différentes et d'acquérir les informations les plus pertinentes relatives aux traditions culturelles et à l'utilisation du sorgho. Ceci devrait permettre d'assurer la perpétuité et l'utilisation des ressources génétiques du sorgho.

Deux importantes collections *ex-situ* des ressources génétiques du sorgho sont maintenues à l'IAC (Center Asiatique de l'ICRISAT) à Patancheru, Inde et au NPGS à Fort Collins, USA.

##### 1) La collection ICRISAT (Inde)

La collection de sorgho a été initiée dans les années soixante par la fondation Rockefeller en tant que partie du programme indien de recherche en agriculture (Prasada Rao et Gopal Reddy, 1989). A partir de 1976, l'ICRISAT a été mandaté par l'IBPGR (devenu IPGRI) pour prendre en charge la collection alors existante et pour poursuivre les collectes de sorgho (Prasada Rao et al., 1995). De fait, l'ICRISAT s'engage à mettre à la disposition de tout utilisateur (scientifiques et agriculteurs) 80% des accessions de sorgho acquises avant la convention sur la biodiversité (Eberhart et al., 1998).

Un total de 36 719 accessions dont 30 950 variétés traditionnelles, 5 327 lignées améliorées et 417 formes sauvages sont conservées dans des conditions de moyen terme à 4°C et 20% d'humidité relative. Plus de 51% d'entre elles sont également maintenues dans des conditions de long terme à -20°C. Les accessions dont le seuil de germination est inférieur à 85% ou pour lesquelles le stock a atteint la limite de 5grs de graines, sont régénérées. Cette étape est conduite en conditions contrôlées d'autofécondation dans les champs expérimentaux de l'IAC. Une trentaine de panicules sont récoltées. Les graines en mélange fournissent la réserve des semences (Stenhouse et al., 1997).

A l'IAC, les évaluations pour 22 caractères morphologiques sont réalisées régulièrement et les données sont disponibles et complètes pour plus de 98% de la collection



**Tableau 5** : évaluation du sorgho à l'IAC  
(d'après Stenhouse et al., 1997) ; caractères et  
nombre correspondant d'accessions évaluées.

caractère	nombre d'accessions évaluées
moisissure du grain	7 132
anthracnose	13 000
mildiou	13 000
rouille	5 000
striga	14 000
<i>Atherigona</i>	31 000
foreur de la tige	30 000
cécidomyie	15 000
punaie des panicules	15 000
puceron vert	28 000
contenu protéinique	11 000
teneur en lysine	9 900
teneur en sucre de la tige	9 000

totale. Cependant, l'évaluation des caractères d'intérêt (notamment des résistances à des pathogènes, qualité du grain ou encore rendement en grain) ne concernent qu'une portion de la collection totale (Stenhouse et al., 1997) (Tableau 5).

## 2) La collection NPGS (USA)

Une grande partie de la collection de l'ICRISAT est dupliquée aux USA. Quelques pays non représentés dans la collection ICRISAT, le sont dans la collection américaine. C'est le cas de l'Afrique équatoriale, la Guinée, le Libéria, la Libye, le Kazakhstan, le territoire d'Oman. La mission du NPGS est de collecter, préserver, évaluer, améliorer et distribuer les ressources génétiques des céréales économiquement importantes (Eberhart et al., 1998). Le premier cycle de croissance des accessions est assuré à St Croix dans les Iles Vierges (USA) sous les conditions régulées de quarantaine. La conservation de la collection de base est maintenue sur du long terme à Fort Collins (Colorado, USA). La majorité des accessions est conservée de façon conventionnelle à -18°C, alors qu'une faible portion est maintenue par cryoconservation dans l'azote liquide à -196°C. Une collection active représentant un sous-ensemble de la collection, est conservée à Griffin (Georgia, USA). Les semences sont régénérées à Mayagüez, Puerto Rico, et les stock génétiques US sont maintenus à la station expérimentale de l'Université Texas A&M.

Pour seulement les deux tiers de la collection (31 708 accessions), le pays d'origine est connu. L'évaluation de la collection totale est en cours ; en 1995, environ 13% des 42 221 accessions recensées dans la collection américaine étaient complètement évaluées (Dalhberg et Spinks, 1995).

Finalement, après la prise de conscience de la nécessité de collecter (avant qu'il ne soit trop tard) les ressources génétiques, une autre étape a été franchie en cette fin de siècle, celle de la nécessité d'évaluer les ressources génétiques faute de transformer les banques de gènes en collections poussiéreuses et inutilisables. Aujourd'hui, des données sont disponibles, leur analyse est nécessaire pour fournir des éléments de réponse aux questions posées par la gestion et l'utilisation de collections de grande taille.

## C - OBJECTIFS DU TRAVAIL

Le sorgho est une plante aujourd'hui cultivée sur tous les continents. La domestication du sorgho réalisée dans le quart nord-est de l'Afrique s'accompagne de plusieurs centres de diversification secondaire. L'ICRISAT a pour mandat de collecter et de préserver les ressources génétiques du sorgho. La collection mise en place à l'IAC répond-elle effectivement à son statut de collection mondiale ?

La collection de l'ICRISAT s'accompagne d'une base de données complètes sur plus de 20 000 accessions de variétés traditionnelles acquises sur plusieurs années. Comment peut-on prendre en compte les informations de cette base de données ?

Les marqueurs moléculaires sont accessibles chez le sorgho. Peut-on utiliser de tels marqueurs pour faciliter la gestion et l'utilisation des ressources génétiques du sorgho ?

La collection de sorgho comporte aujourd'hui plus de 35 000 accessions. Est-il possible de définir une ou des stratégies facilitant la gestion et l'utilisation de cette collection ?

Dans le cadre de ce travail, nous nous intéresserons exclusivement aux ressources génétiques du sorgho et notamment à la collection de l'ICRISAT en Inde. Cette collection comporte 36 719 accessions de sorghos acquises dans plus de 90 pays très largement répartis à travers le monde

L'objectif de ce travail est de tenter d'apporter des éléments de réponse à trois grandes questions :

### 1) Comment peut-on faciliter l'accès à la collection des ressources génétiques du sorgho ?

Dans cette première partie, nous nous proposons tout d'abord de montrer que les données d'évaluation enregistrées dans la base sont exploitables bien que les conditions d'évaluation au cours des années n'aient pas été strictement semblables. Puis nous aborderons l'étude descriptive détaillée de la collection de base. Pour ce faire, nous étudierons tout d'abord la diversité géographique, raciale et morpho-agronomique (chapitre 1). Enfin dans le chapitre 2, nous analyserons l'organisation de la diversité des sorghos.

### 2) Quelle stratégie utiliser pour constituer des core collections ?



Dans cette deuxième partie, différentes stratégies d'échantillonnage seront comparées. A partir de la collection stratifiée, trois prélèvements aléatoires seront établis et comparés à la collection de base (chapitre 1). Par la suite, un des échantillonnages stratifiés aléatoires réduit à 1% de la taille de la collection de base sera comparé à deux échantillonnages non aléatoires ; ces trois sous-ensembles étant à leur tour comparés à la collection de base (chapitre 2). L'impact de la stratégie de prélèvement est examiné sur des caractères morpho-agronomiques non directement pris en compte lors de l'échantillonnage.

3) Quel est l'impact d'un prélèvement basé sur un jeu de caractères sur la diversité estimée à partir d'un autre jeu de caractères ?

Pour cette troisième partie, nous évaluerons la diversité moléculaire retenue après utilisation de différentes stratégies d'échantillonnage basées sur les caractères morpho-agronomiques (chapitre 1). D'éventuelles associations seront également recherchées entre marqueurs moléculaires, caractères morpho-agronomiques, raciaux et origines géographiques (chapitre 2). Enfin, nous étudierons l'impact sur la diversité morpho-agronomique d'un échantillonnage effectué à partir des caractères moléculaires (chapitre 3).

## **1<sup>ÈRE</sup> PARTIE - ETUDE DE LA COLLECTION DE BASE**

**Tableau I.1 : liste des descripteurs morpho-agronomiques disponibles dans la base de données de l'ICRISAT, et ceux retenus dans notre étude (Chp 1)**

caractères	code	modalités définies par l'IPGRI (IBPGR/ICRISAT, 1993)	modalités retenues dans notre étude	déterminisme génétique
date de floraison de 50% des plantes (jours depuis la date de semis) ➤ saison kharif	FLK		<ul style="list-style-type: none"> <li>• précoce (&lt;70)</li> <li>• tardive (&gt;130)</li> </ul>	
➤ saison rabi	FLR			
nombre des talles racinaires	BT			
hauteur de plantes (cm) ➤ saison kharif	PHTK		<ul style="list-style-type: none"> <li>• petite (&lt;150)</li> <li>• grande (&gt;450)</li> </ul>	
➤ saison rabi	PHTK			
longueur du pédoncule (cm)	PEDEX		<ul style="list-style-type: none"> <li>• nulle (=0)</li> <li>• long (&gt;30)</li> </ul>	
longueur de la panicule (cm)	EHLG			
largeur de la panicule (cm)	EHWD			
taille des grains (mm)	GRS			
poids de 100 grains (g)	SWT		<ul style="list-style-type: none"> <li>• léger (&lt;2grs)</li> <li>• lourd (&gt;4grs)</li> </ul>	
couleur de la plante	PIG	<ul style="list-style-type: none"> <li>• pigmentée</li> <li>• tan</li> </ul>	<ul style="list-style-type: none"> <li>• pigmentée</li> <li>• tan</li> </ul>	
couleur de la nervure médiane	MRC	<ul style="list-style-type: none"> <li>• blanche</li> <li>• verdâtre</li> <li>• jaune</li> <li>• brune</li> </ul>	<ul style="list-style-type: none"> <li>• blanche</li> <li>• verdâtre</li> <li>• jaune</li> <li>• brune</li> </ul>	
couleur des glumes	GLC	<ul style="list-style-type: none"> <li>• blanc • paille • jaune • gris</li> <li>• brun • brunâtre • partiellement brun</li> <li>• pourpre • partiellement pourpre • noir</li> <li>• rouge • rougeâtre • rouge/brun</li> </ul>	<ul style="list-style-type: none"> <li>• sienne</li> <li>• mahogany</li> <li>• noir et pourpre</li> <li>• rouge</li> </ul>	pQ ou pq' pq Pq PQ ou Pq'
compacité et forme de la panicule	EHCS	<ul style="list-style-type: none"> <li>• semi compacte elliptique • compacte elliptique • semi compacte ovale • compacte ovale</li> <li>• semi lâche et ramifications l<sup>aires</sup> érigées • lâche et ramifications l<sup>aires</sup> érigées</li> <li>• très lâche et ramifications l<sup>aires</sup> érigées</li> <li>• semi lâche et ramifications l<sup>aires</sup> retombantes • lâche et ramifications l<sup>aires</sup> retombantes</li> <li>• très lâche et ramifications l<sup>aires</sup> retombantes</li> </ul>	<ul style="list-style-type: none"> <li>• compacte</li> <li>• lâche/érigées</li> <li>• lâche/tombantes</li> </ul>	
couleur des grains	GRC	<ul style="list-style-type: none"> <li>• brun • brunâtre • rouge/brun</li> <li>• blanc crayeux • gris</li> <li>• rouge • rougeâtre • noir • pourpre</li> <li>• blanc</li> <li>• jaune</li> <li>• paille</li> <li>• blanc et rouge</li> </ul>	<ul style="list-style-type: none"> <li>• brun</li> <li>• gris</li> <li>• rouge</li> <li>• blanc</li> <li>• jaune</li> <li>• paille</li> <li>• "xenia"</li> </ul>	<u>sous-couche / diffuseur / gène de coloration</u> P / P / Ry ou RY P / A / Ry P ou A / A / RY P ou A / A / Ry P ou A / A / rY P ou A / A / ry P ou A / A / "xenia"
couverture du grain par les glumes	COV	de • 1 : non couvert à • 5 : couverture complète	<ul style="list-style-type: none"> <li>• 1 : non-couvert</li> <li>• 5 : recouvert</li> </ul>	
texture de l'albumen	COR	de • 1 : grain vitreux à • 5 : grain farineux	<ul style="list-style-type: none"> <li>• 1 : vitreux</li> <li>• 5 : farineux</li> </ul>	
lustrosité des grains	LUS	<ul style="list-style-type: none"> <li>• lustré</li> <li>• non-lustré</li> </ul>		
sous-couche brune	SC	<ul style="list-style-type: none"> <li>• présence</li> <li>• absence</li> </ul>	<ul style="list-style-type: none"> <li>• présence</li> <li>• absence</li> </ul>	



## 1<sup>ÈRE</sup> PARTIE - ÉTUDE DE LA COLLECTION DE BASE

La base de données de l'ICRISAT est constituée depuis 1972 à partir de caractérisations morpho-agronomiques effectuées sur vertisols à Patancheru (Inde) i.e. 17°25' latitude nord, 78° longitude est. Les dispositifs expérimentaux sont généralement conduits pour les deux saisons caractéristiques des régions tropicales semi-arides; la saison des pluies ("kharif") et la saison sèche ("rabi") (Appa Rao et al., 1996). La saison kharif se caractérise par des jours longs et des conditions climatiques chaudes et humides. A l'inverse, la saison rabi correspond à des jours courts et des conditions sèches avec des températures plus basses. Les informations accessibles dans la base de données de l'ICRISAT concernent les origines géographiques, la classification raciale et des données de caractérisation morpho-agronomique. Dix sept descripteurs morpho-agronomiques, appartenant à la liste recommandée par l'IPGRI, (IBPGR/ICRISAT, 1993) sont étudiés pendant la saison rabi. Seuls deux d'entre eux (date de floraison et hauteur de plantes) sont également notés à la saison kharif (Tableau I.1).

Les données acquises régulièrement au cours des années, soulèvent la question de leur utilisation effective étant données les conditions de leur acquisition. En effet, jusqu'à présent, les essais ont été conduits en l'absence de témoins. Les dates de semis et de récolte, les données climatiques ne figurent pas dans la base de données alors que bon nombre des caractères soumis à l'évaluation sont considérés fortement sensibles aux conditions environnementales. Compte tenu de ces réserves, il nous semble nécessaire de montrer que les valeurs enregistrées dans la base de données toutes années confondues peuvent être tout de même utilisées pour des analyses ultérieures.

Nous avons donc tout d'abord entrepris l'évaluation de 642 accessions dans les conditions d'expérimentation décrites ci-après.

Les 642 accessions représentent les 15 races de sorgho et couvrent une large gamme d'origines géographiques (43 pays). Une variété standard (IS 1054) bien adaptée aux conditions locales est utilisée comme témoin dans nos dispositifs expérimentaux. Les essais ont été conduits à l'IAC, sur vertisol, entre 1996 et 1998. Un dispositif expérimental aléatoire est établi pour les saisons rabi et kharif. A partir de ce témoin régulièrement disposé dans chaque parcelle (1 rang de témoin intercalé tous les 20 rangs), les effets environnementaux au sein du champ sont estimés par une analyse de

\* : corrélations significatives au seuil de 5%

variance d'ordre 1 (ANOVA) (Statsoft, 1995). Les caractères concernés sont la date de floraison et la hauteur de plante évalués au cours des deux saisons, puis l'exertion pédonculaire, la longueur et la largeur de la panicule, la taille du grain et le poids de 100 grains estimés uniquement en saison rabi. Il s'agit alors, pour chacun de ces neuf caractères, de comparer les variances et si possible, les moyennes des témoins entre les différentes années. Un test d'homogénéité de variance de Levene a permis d'évaluer les différences de dispersion entre groupes en comparant les variances d'un caractère. Lorsque les variances sont homogènes entre deux groupes, les moyennes du caractère sont comparées par un test de Newman-Keuls.

Les résultats obtenus montrent que les variables de date de floraison en rabi (FLR), longueur de la panicule (EHLG) et hauteur de la plante en kharif (PHTK) ont des variances significativement différentes (test de Levene avec  $p < 0,05$ ) (Tableau I.2). Pour les caractères de hauteur de plante à la saison rabi (PHTR), la longueur du pédoncule (PEDEX), la largeur de la panicule (EHWD) et la date de floraison à la saison kharif (FLK), bien que les variances soient homogènes, les moyennes des plantes témoins sont significativement différentes entre les essais (test de Newman-Keuls significatif à  $p < 0,05$ ). De façon à limiter les effets dus au dispositif expérimental, nous avons procédé à une standardisation de la manière suivante. Chaque parcelle est découpée en surfaces contenant de 3 à 5 témoins. Pour chaque caractère quantitatif et pour chaque dispositif expérimental, la moyenne obtenue pour les témoins est retranchée aux valeurs enregistrées pour chaque accession.

Puis, nous avons réalisé une étude des corrélations pour les 642 accessions, entre les nouvelles données morpho-agronomiques standardisées, et les valeurs prises dans la base de données (Tableau I.3). Toutes les corrélations sont significatives au seuil de 5%. De très fortes corrélations sont obtenues pour des caractères tels que la date de floraison à la saison kharif, la hauteur de la plante pour les deux saisons, les caractéristiques de la panicule ou encore le poids des grains.

Il semble donc que malgré les réserves apportées sur les dispositifs expérimentaux, on puisse considérer que, de manière globale, les données enregistrées dans la base sont cohérentes. Elles peuvent donc être utilisées comme telles pour l'étude de la collection de base.

Pour estimer la diversité phénotypique de la collection, nous nous sommes intéressés aux distributions des accessions dans les différentes modalités des



### Encadré 1 : indice de diversité de Shannon-Weaver

Originellement utilisé en écologie pour évaluer la diversité des espèces dans des communautés, l'indice de Shannon-Weaver peut également être appliqué à l'estimation de la diversité d'un caractère à partir des classes phénotypiques.

$$H_c' = - \sum_{i=1}^n p_i \log_e p_i \quad (1),$$

avec pour un caractère donné C

- $n$  : le nombre de classes phénotypiques
- $p_i$  : la proportion de cas dans la  $i^{\text{ème}}$  classe

L'indice de Shannon-Weaver présente la propriété intéressante d'additivité, ce qui permet d'estimer la variance de l'indice de diversité (Pielou, 1969, cité dans Jain et al., 1974). Hutcheson (1970) propose un test pour la comparaison des diversités basées sur l'indice de Shannon-Weaver.

Le calcul de la variance est estimé par

$$\text{Var}(H') = [\sum p_i \log_e^2 p_i - (\sum p_i \log_e p_i)^2]/N + (n-1)/2N^2$$

avec  $N$  : nombre d'observations.

Une comparaison d'indices de diversité peut alors être effectuée par un test-t :

$$T = (H'_1 - H'_2) / [\text{Var}(H'_1) + \text{Var}(H'_2)]^{1/2}$$

avec pour degré de liberté : d.d.l. =  $[\text{Var}(H_1) + \text{Var}(H_2)] / [\text{Var}(H_1)^2/N_1 + \text{Var}(H_2)^2/N_2]$ ,  
où  $N_1$  et  $N_2$  sont les nombres de cas dans les deux groupes.

De nombreuses études ont utilisé ces propriétés d'additivité pour estimer la variance de la diversité en fonction de la distribution géographique (Jain et al., 1974 ; Wu et Jain, 1977 ; Tolbert et al., 1978 ; Bogyo et al., 1980 ; Weltzien, 1989 ; Perry et McIntosh, 1991 ; Galwey, 1995 ; Li et al., 1996; Balfourier et al, 1998; Bisht et al., 1998 ; Polignano et al., 1999).

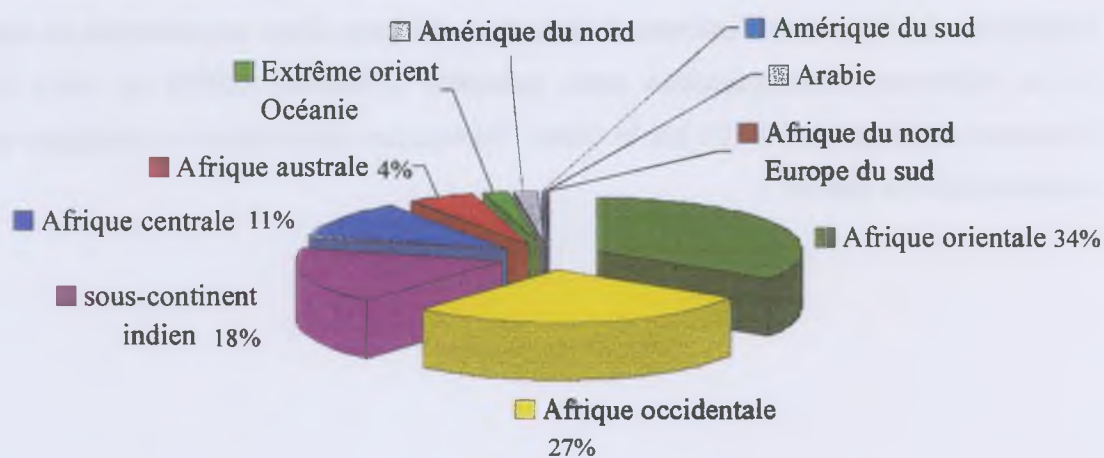
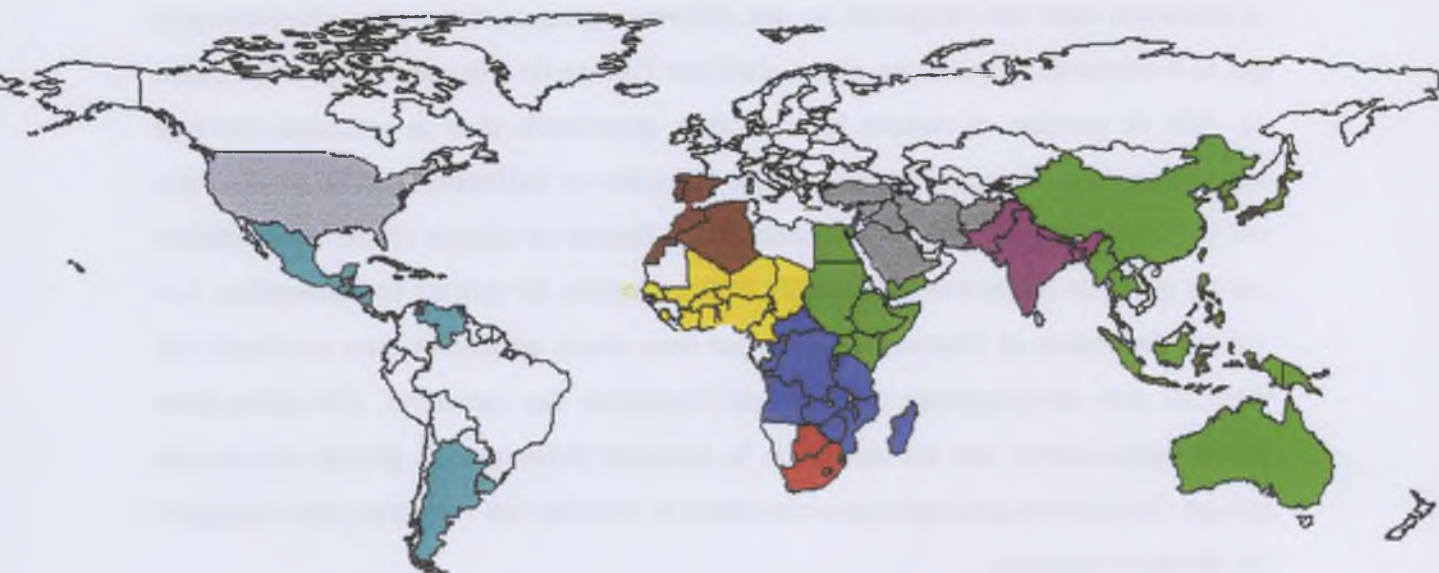
Mahajan et al., (1996 a et b), d'après les travaux de Galwey, (1992 cités par les précédents auteurs) proposent la constitution de core collection à partir de l'estimation de l'indice de Shannon-Weaver calculé sur les caractères qualitatifs importants.

L'indice de Shannon-Weaver initialement appliqué aux variables qualitatives, s'adapte également aux variables quantitatives, après découpage de ces dernières en classes phénotypiques (Holcomb et al., 1977 ; Charmet et Balfourier, 1995).

Parmi les différents indices de diversité génétique utilisables pour les caractères qualitatifs, l'indice de Shannon-Weaver présente une caractéristique intéressante, car il combine à la fois le concept de la richesse allélique et celui de la distribution des fréquences alléliques (van Hintum, 1995).

caractères morpho-agronomiques. Pour mesurer la diversité phénotypique globale de la collection dans son intégralité ou des différents groupes raciaux ou géographiques qui la composent, nous avons choisi d'utiliser l'indice de Shannon-Weaver (Encadré 1). Afin de prendre en compte les caractères quantitatifs dont la variation continue suggère un mode de contrôle polygénique (Charmet et Balfourier, 1995), ces derniers ont été découpés en classe de fréquences. Les bornes de chaque classe sont définies par les quartiles obtenus sur l'ensemble de la collection de variétés traditionnelles. Les valeurs de l'indice de Shannon-Weaver que nous avons présentées dans nos études de diversité sont des moyennes obtenues sur l'ensemble des caractères. Cet indice nous permet ainsi d'avoir une estimation de la diversité phénotypique globale au sein du groupe (racial et/ou géographique) considéré, et d'utiliser les variances pour comparer les diversités relatives.

Cette partie sera composée de deux chapitres ayant chacun fait l'objet d'un article. Le premier chapitre traite de la diversité de la collection (article soumis à *Euphytica*). Le deuxième s'intéresse à la mise en évidence d'une structuration au sein de la collection, (communication orale présentée à Méribel (1999) au cours du séminaire organisé par l'INRA sur le thème "Marqueurs moléculaires et stratégies en amélioration des plantes").



**Figure I-1.1 :** distribution géographique des variétés traditionnelles de sorgho cultivés.



## CHAPITRE 1 : Diversité morpho-agronomique

Une utilisation optimale des ressources génétiques de sorgho passe nécessairement par la connaissance de la diversité phénotypique de la collection. A partir des informations contenues dans la base de données de l'ICRISAT, nous nous proposons de décrire la diversité morpho-agronomique de la collection.

La collection de sorgho étudiée, restreinte d'une part aux seules variétés traditionnelles d'autre part aux variétés disposant des données complètes, comporte un total de 22 473 accessions. En tenant compte du schéma de domestication et de diversification des sorghos (Figure 4 dans l'introduction), 10 grandes régions ont été définies (Figure I-1.1). Douze descripteurs morpho-agronomiques (Tableau I.1) sont étudiés en considérant tout d'abord la collection dans son ensemble, puis les 15 races et les distributions géographiques séparément.

Nos analyses montrent que la distribution géographique des accessions acquises par l'ICRISAT, est inégalement répartie sur les 10 régions. Cette observation semble toutefois cohérente avec :

- une production de sorgho principalement dans les zones tropicales semi-arides pour lesquelles il est bien adapté. Ainsi l'Afrique orientale, centrale et occidentale contribuent pour 73% alors que les régions Nord Afrique plus Sud Europe et Arabie fournissent à elles deux moins de 0,5% des accessions ;
- le schéma de domestication du sorgho. Le centre d'origine des sorghos contribue pour 34% des accessions.

Cependant, des pays voire des régions sont mal représentés et notamment la Chine, un des centres de diversification secondaire des sorghos. En effet, les variétés chinoises correspondent à 1,2% de la collection étudiée alors que plus de 10 000 accessions sont conservées dans la banque de gènes de Pékin (Deu et al., 1999).

Face à cette répartition géographique quelque peu déséquilibrée, nous avons choisi de considérer uniquement les régions dont la représentation pouvait être à priori satisfaisante. Ainsi, seules les régions contribuant pour au moins 5% des accessions ont été plus finement analysées. Elles sont au nombre de cinq et correspondent aux différentes régions africaines, exception faite de la région Nord, et au sous continent indien (Tableau I-1.1).

**Tableau I-1.1** : caractérisation phénotypique des régions. Distribution (en pourcentage d'accessions) dans différentes modalités des caractères morpho-agronomiques pour la collection des variétés traditionnelles (total) et pour chacune des cinq régions principales.  
 † : seules les modalités les plus extrêmes sont considérées.

caractère	modalité	total	Afrique orientale	Afrique occidentale	Sous-continent indien	Afrique centrale	Afrique australe
FLK†	précoce	22	22	3	18	22	89
	tardive	19	14	36	9	20	≡0
PHTK†	petite	2	1	1	1	3	9
	grande	9	6	15	8	12	≡0
PEDEX†	nul	6	5	4	9	8	6
	long	7	10	6	3	3	2
SWT†	léger	12	11	6	16	25	7
	lourd	17	20	30	9	1	2
PIG	pigmentée	98	97	99	98	95	100
	tan	2	3	1	2	5	≡0
MRC	blanche	82	84	93	75	77	65
	verdâtre	17	15	7	25	22	35
	jaune	1	1	≡0	≡0	≡0	≡0
	brune	≡0	≡0	-	-	≡0	-
GLC	sienne	31	25	16	60	29	52
	acajou	11	11	11	11	12	6
	noir et pourpre	42	49	53	19	39	28
	rouge	16	15	20	10	20	14
COV†	non-couvert	5	≡0	10	6	8	1
	recouvert	3	2	1	4	1	2
EHCS	compacte	45	50	39	68	25	32
	lâche/érigées	47	45	52	24	61	65
	lâche/tombantes	8	6	10	7	14	3
COR†	vitreux	4	≡0	8	4	7	1
	farineux	14	23	10	3	18	3
SC	présence	34	47	31	10	33	27
	absence	66	53	69	90	67	73
GRC	brun	29	39	19	10	41	23
	gris	10	19	8	2	5	3
	rouge	16	16	21	6	9	33
	blanc	20	7	26	35	22	24
	jaune	5	6	7	8	1	≡0
	paille	20	13	18	40	22	18
	"xenia"	≡0	≡0	≡0	-	-	-

La plus grande diversité de phénotypes est rencontrée en Afrique orientale. Néanmoins, certains d'entre eux sont relativement rares. Peu de variétés locales ont des grains non-couverts par les glumes et/ou un albumen vitreux. Dans cette région, 22% des sorgho sont des variétés précoces, 39% ont des grains bruns. Par contre, 59% des variétés à nervure centrale jaune, 64% des variétés à grain blanc crayeux, 57% des variétés à grains farineux et près de 55% des variétés à long pédoncule sont originaires de l'Afrique orientale.

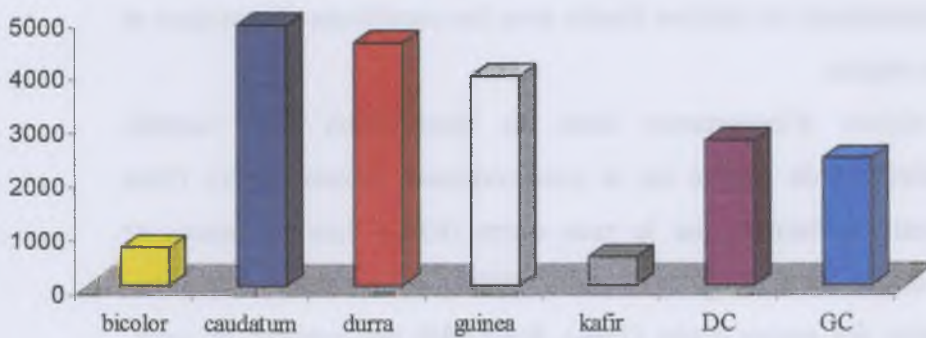
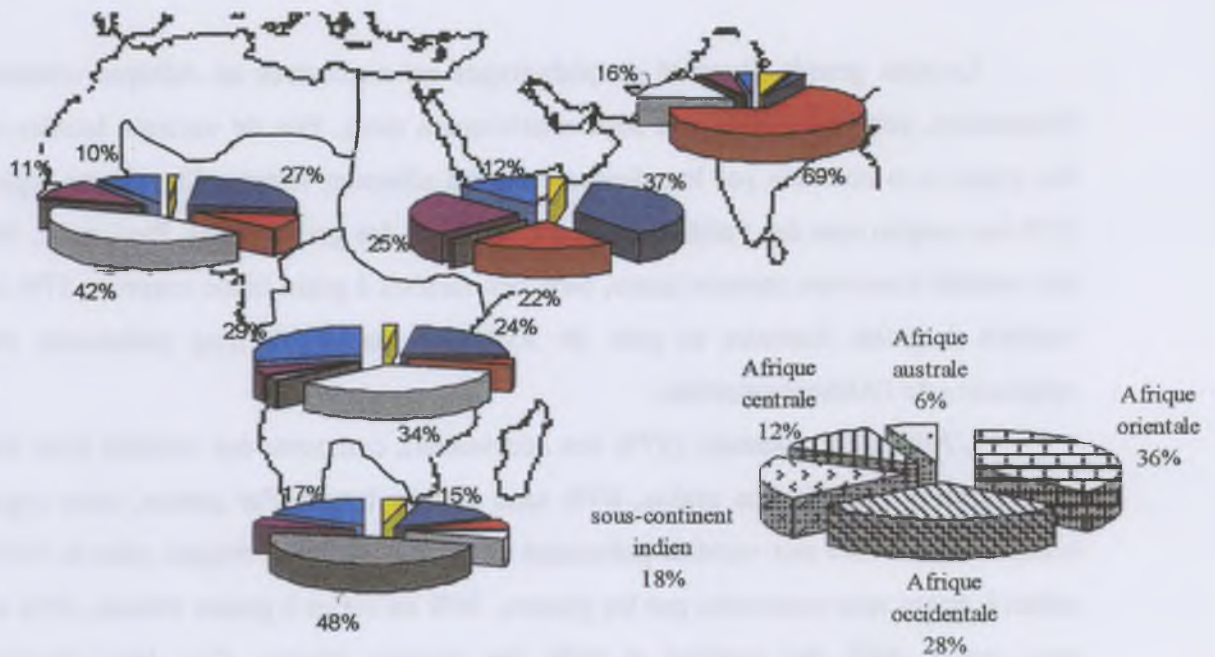
L'Afrique occidentale, (27% des accessions), comporte des variétés dont 36% sont tardives, 30% à gros grains, 69% sans couche brune. Par contre, cette région contribue pour 75% aux variétés présentant des grains blancs et rouges, plus de 56% à celles à grains non recouverts par les glumes, 56% de celles à grains vitreux, 48% des gros grains, 66% des tardives et 46% des grandes plantes. Ces deux dernières caractéristiques sont certainement en relation étroite avec les conditions climatiques et de photopériode de cette région.

La troisième région d'importance dans la distribution des variétés traditionnelles de la collection de sorgho est le sous-continent indien (18%). Cette région est principalement représentée par la race durra (63%). Les tendances se dessinent en faveur de variétés avec des glumes claires (60%), et/ou l'absence de sous-couche brune (90%), et/ou des grains clairs (75%). Pour 68% des variétés du sous-continent indien, les panicules sont compactes. Cette caractéristique est associée à une bonne adaptation à la sécheresse. Dix huit pour-cent des variétés acquises dans cette région sont précoces mais seul 1% sont de petites tailles. Cette observation est en contradiction avec la corrélation positive ( $r = 0,65$ ) notée entre précocité et petite hauteur de plante.

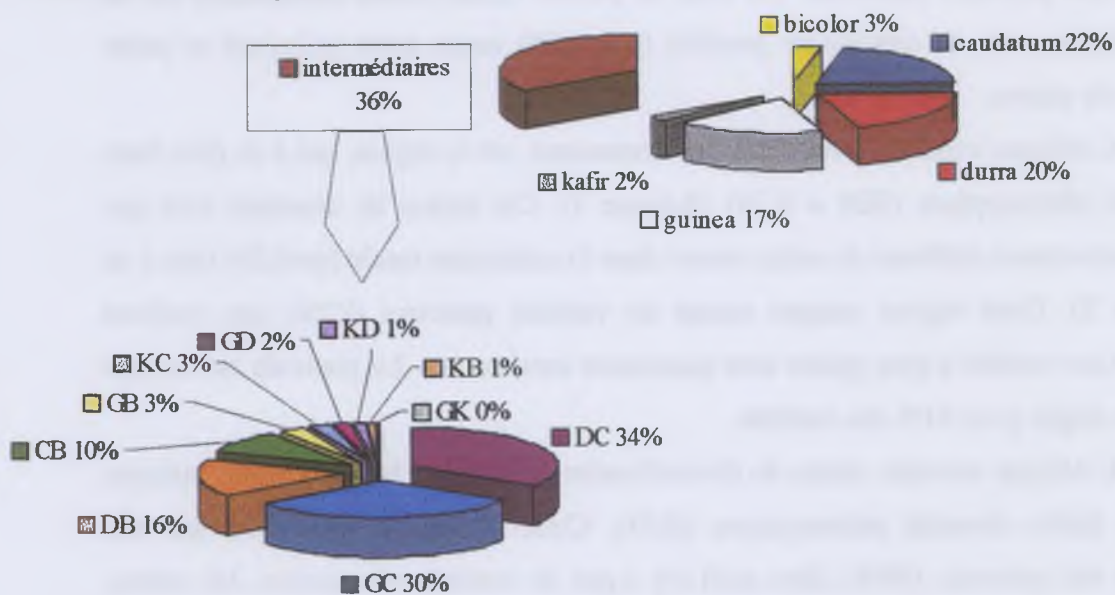
L'Afrique centrale, avec 11% des accessions, est la région qui a la plus forte diversité phénotypique ( $SDI = 0,74$ ) (Annexe 1). Cet indice de diversité n'est pas significativement différent de celui trouvé dans la collection totale ( $p > 0,24$ ) (test-t en Annexe 2). Cette région compte autant de variétés précoces (22%) que tardives (20%). Les variétés à gros grains sont quasiment inexistantes. La panicule est de type lâche et érigée pour 61% des variétés.

L'Afrique australe, centre de diversification secondaire le plus récent, présente la plus faible diversité phénotypique (0,54). Cette région se caractérise par des variétés très précoces (89%) alors qu'il n'y a pas de variétés très tardive. De même, 9% des variétés donnent de très petites plantes alors qu'aucune ne conduit à de très grandes plantes. La majorité des variétés ont une panicule lâche et érigée (65%).





**Figure I-1.2 :** distribution géographique des variétés traditionnelles de sorgho de la collection de l'ICRISAT. Ne sont représentées que les 5 races principales, les 2 races intermédiaires les plus fréquentes et les 5 régions majeures.



**Figure I-1.3 :** distribution raciale des variétés traditionnelles de sorgho cultivés, d'après la classification raciale proposée par Harlan et de Wet (1972).

La distribution des races est différente entre les cinq régions étudiées (Figure I-1.2). Toutes les races sont représentées dans la région d'Afrique orientale, et trois d'entre elles sont largement majoritaires ; caudatum, DC et durra par ordre d'importance. Les sorghos originaires d'Afrique occidentale sont essentiellement des variétés guinea et caudatum. Dans le sous-continent indien, la majorité des sorghos collectés sont des variétés de race durra. Guinea, GC et caudatum sont les principales races présentes dans les accessions de la région d'Afrique centrale. En Afrique australe, la race kafir est prédominante et représente près de la moitié des variétés collectées.

La distribution raciale dans la collection est aussi déséquilibrée (Figure 1-1.3). Les races caudatum, durra, guinea, DC et GC représentent à elles seules 82% des accessions. En ne considérant que les cinq races majeures et les deux races intermédiaires fortement représentées dans la collection (DC et GC), on constate que la diversité phénotypique intra-race est variable. Certaines races montrent un indice de diversité plutôt faible malgré un effectif élevé sans doute en raison d'intérêts agronomiques spécifiques, et d'une utilisation relativement ciblée des variétés cultivées. Par exemple, l'indice de Shannon-Weaver est de 0,56 pour la race durra (4 471 accessions) contre 0,68 pour bicolor (521 accessions).

La race caudatum est largement représentée dans la collection (22%) et son indice de diversité est relativement élevé (0,67). Cette race comprend des variétés à très petits grains (13%) et des variétés à très gros grains (17%) (Tableau I-1.2). La présence d'une sous-couche brune est une tendance fortement marquée (70%), et pour la majorité des variétés caudatum les grains sont bruns (61%). Cette caractéristique s'accompagne d'une teneur élevée en tanin ce qui confère un goût amer aux grains de telles variétés et par conséquent une résistance à l'attaque des oiseaux, des insectes, des moisissures (Fliedel et al., 1996). Ces caractéristiques s'accompagnent de la réduction de la qualité alimentaire de ces sorghos bruns. Cependant, de telles variétés peuvent être utilisées dans certaines préparations culinaires, ou encore pour la fabrication de la bière opaque de sorgho (Fliedel et al., 1996). Pour un bon nombre des variétés, l'albumen est farineux (35%) ce qui permet à cette race d'avoir un bon rendement en amidon, mais un faible rendement au décorticage et une pureté des produits peu satisfaisante. Néanmoins, quelques variétés ont un albumen vitreux. Une bonne représentation des différentes formes de la panicule permet la culture des variétés caudatum dans les zones sèches aussi bien qu'humides. Au sein de la race caudatum



**Tableau I-1.2** : caractérisation phénotypique des races. Distribution (en pourcentage) des accessions dans différentes modalités des caractères morpho-agronomiques pour la collection des variétés traditionnelles (total), pour chacune des cinq races principales et pour les deux races intermédiaires fortement représentées dans la collection.

† : seules les modalités les plus extrêmes sont considérées.

caractère	modalité	total	bicolor	caudatum	durra	guinea	kafir	DC	GC
FLK†	précoce	22	40	13	15	4	94	39	26
	tardive	19	12	14	10	42	-	13	14
PHTK†	petite	2	1	2	1	≡0	8	3	5
	grande	9	7	2	4	27	-	3	6
PEDEX†	nul	6	4	5	11	3	4	5	6
	long	7	10	5	4	9	2	13	5
SWT†	léger	12	50	13	5	20	2	4	8
	lourd	17	2	17	24	8	2	34	10
PIG	pigmentée	98	97	98	99	97	99	99	93
	tan	2	3	2	1	3	1	1	7
MRC	blanche	82	79	81	76	97	63	90	70
	verdâtre	17	19	18	22	3	37	9	29
	jaune	1	2	1	2	≡0	≡0	1	1
	brune	≡0	-	≡0	≡0	≡0	-	≡0	≡0
GLC	sienne	31	17	10	56	26	57	28	21
	acajou	11	14	6	10	14	6	13	11
	noir et pourpre	42	54	59	26	43	24	49	44
	rouge	16	15	25	8	17	13	10	24
COV†	non-couvert	5	-	≡0	-	28	-	-	≡0
	recouvert	3	59	≡0	≡0	≡0	-	≡0	-
EHCS	compacte	45	10	58	86	2	38	55	30
	lâche/érigées	47	71	36	14	75	61	44	63
	lâche/tombantes	8	19	6	1	24	1	1	7
COR†	vitreux	4	6	≡0	≡0	19	1	≡0	1
	farineux	14	2	35	6	1	3	18	13
SC	presence	34	50	69	4	9	13	44	52
	absence	66	50	31	96	91	87	56	48
GRC	brun	29	76	61	5	5	6	17	39
	gris	10	0	20	1	3	2	24	14
	rouge	16	7	9	16	19	50	21	13
	blanc	20	8	6	20	47	24	13	17
	jaune	5	1	1	15	1	≡0	12	1
	paille	20	8	3	43	25	19	13	16
	"xenia"	≡0	-	≡0	≡0	≡0	-	≡0	-



on trouve autant de variétés précoces (13%) que tardives (14%). Cette caractéristique confère à la race caudatum un très grand intérêt pour les programmes de sélection.

La race durra représente 20% de la collection. Elle se caractérise par des grains très gros (1/4 des variétés), l'absence de sous-couche brune (96%) et des grains et/ou des glumes essentiellement clairs (respectivement 78% et 56%). Les variétés à grains vitreux sont très rares de même que celles à grains très farineux sont minoritaires (6%). La panicule compacte (86% des cas), rendant difficile le séchage des grains, et la faible exsertion paniculaire (40% des variétés ont un pédoncule de longueur <10 cm dont 11% ont une exsertion nulle) limitant le séchage de la panicule, corroborent le fait que cette race soit adaptée préférentiellement aux régions sèches. Cependant, on note la présence de rares variétés à panicule lâche qui pourraient être mieux adaptées à des régions plus humides. Cette race semble présenter une large gamme de réactions à la photopériode ; 15% des variétés sont très précoces et 10% très tardives. Cette race est très importante dans le sous-continent indien, où elle représente 63% des sorghos cultivés.

La race guinea est bien représentée dans la collection (17%). Elle comporte essentiellement des variétés sans sous-couche brune (91%). Dix neuf pour-cent des variétés ont des grains vitreux. Cette caractéristique est un des critères importants de qualité des grains. Elle confère aux guinea un intérêt alimentaire majeur dans certaines régions et rend compte d'une large participation dans les programmes de sélection en Afrique occidentale. Les variétés guinea ont majoritairement une panicule lâche (98%) et/ou des grains peu ou pas recouverts (28%). Ces caractéristiques sont associées à une bonne aptitude au séchage et valent à cette race d'être préférentiellement cultivée en conditions humides. Les variétés guinea sont essentiellement tardives (42%) et fournissent de grandes plantes (27%). A côté de ces caractéristiques générales de la race (en accord avec les descriptions taxonomiques), certaines variétés sont originales par leur précocité et/ou présentent des grains recouverts par les glumes et/ou des panicules compactes et/ou un albumen farineux. Les caractéristiques majeures des guinea expliquent sa distribution dans les zones équatoriales et intertropicales. Guinea est la race majeure en Afrique occidentale et centrale, où elle représente respectivement 41% et 34% des sorghos cultivés. Sa rare présence en Afrique australe est due à celle de variétés précoces et de hauteur de plante moyenne.

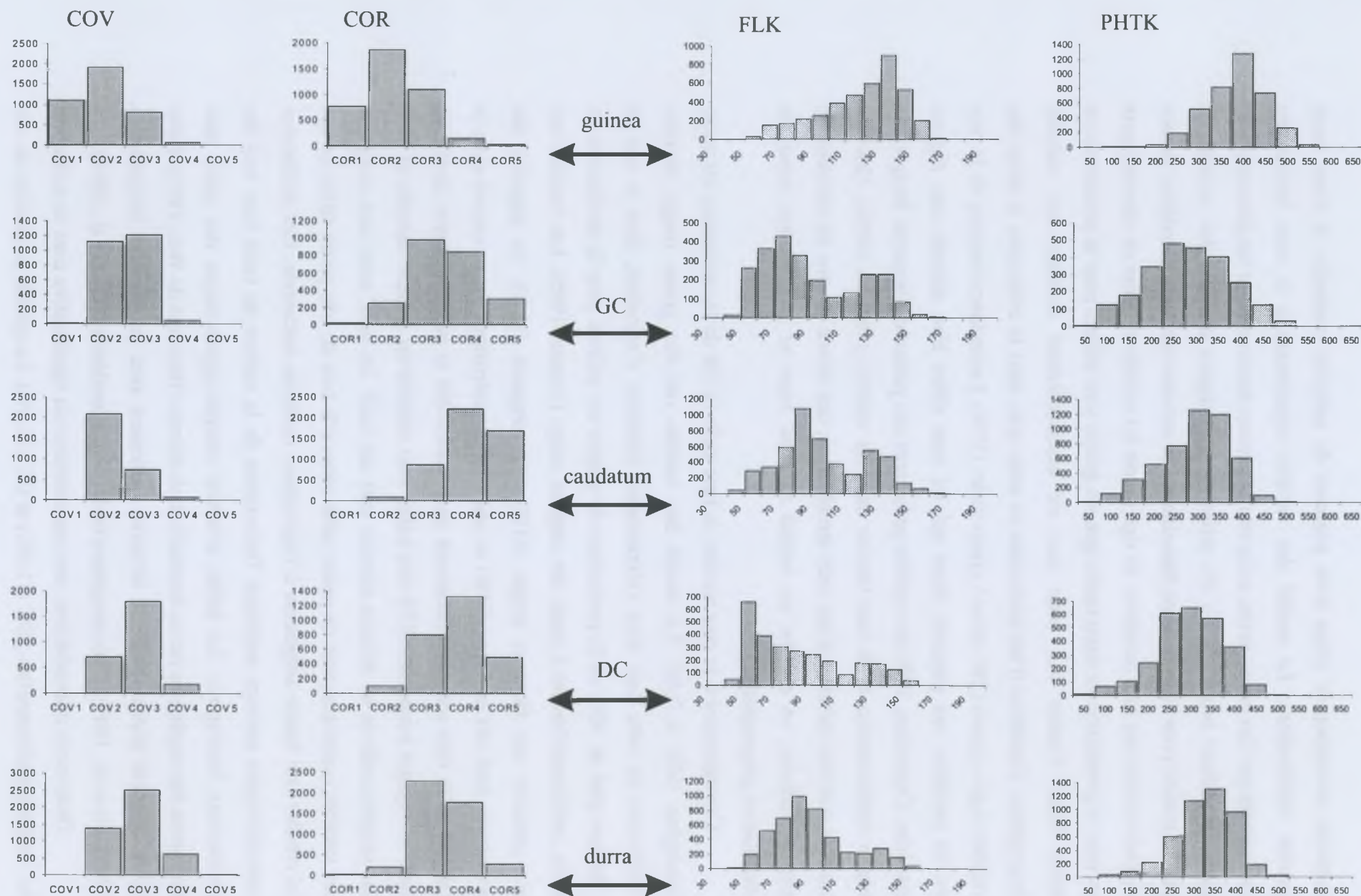
La race bicolor représente 3,2% de la collection. Sa grande variabilité phénotypique (SDI = 0,68) résulte sans doute d'une faible pression de sélection

humaine accompagnée d'une forte pression de sélection naturelle et d'une large gamme d'utilisations. La moitié des variétés appartenant à la race bicolor est caractérisée par des grains petits, et/ou très largement recouverts par les glumes (59%) et/ou de couleur brune (76%). De telles caractéristiques donnent des variétés de pauvre qualité pour l'alimentation humaine et ont souvent conduit à considérer la race bicolor comme une race primitive. Sa tige sucrée est utilisée en guise de chewing-gum ou pour la production de sirop tandis que ses grains sont utilisés pour la production de bière. Dans d'autres cas, cette race est essentiellement destinée aux cultures fourragères. Toutefois, il est intéressant de noter que, dans la collection, il existe des variétés à gros grains (2%) et/ou à grains clairs (17%). La tendance majeure de la race est de posséder une panicule lâche qui lui vaut d'être bien adaptée aux régions humides. Cependant, 10% des variétés présentent une panicule compacte. Malgré une bonne représentation de la race bicolor dans les variétés précoces (40%), 12% des variétés sont très tardives. Cette race semble avoir une bonne gamme de variation de photopériodisme, ce qui est en accord avec sa vaste et relativement constante distribution géographique.

Contrairement à la race bicolor, la race kafir (2,3% de la collection) est assez homogène ( $SDI = 0,46$ ). La moitié des variétés ont des grains rouge, limitant l'utilisation de cette race dans l'alimentation humaine. Cependant, dans le sud de l'Afrique plus de 90% de la production de sorgho est utilisée pour la production de bière, préférentiellement à partir des sorghos rouge (Asante, 1995). Les variétés ont une panicule soit lâche et érigée (61%) soit compacte (38%). La majorité des accessions sont très précoces (94%) et une telle caractéristique est en accord avec le fait que cette race soit essentiellement photo-insensible et principalement distribuée dans les régions tempérées (85% des kafirs sont cultivés en Afrique australe et 43% des sorghos cultivés en Afrique australe sont des kafir. De plus, cette race comporte des variétés essentiellement de petite taille (dont 8% sont de très petite taille) ce qui lui confère une bonne adaptation à l'agriculture moderne mécanisée. Ces différentes caractéristiques peuvent expliquer l'orientation de la culture de cette race vers les productions fourragères. La faible diversité morpho-agronomique des kafir peut également être expliquée par sa diversification récente (Harlan et de Wet, 1972). Cette faible diversité phénotypique se retrouve également avec les données isozymiques (Ollitrault et al., 1989) et les marqueurs moléculaires nucléaires (Deu et al., 1994).

Deux races intermédiaires ont une contribution significative dans la collection des variétés traditionnelles ; DC (11,6%) et GC (10,2%). La grande proportion de ces





**Figure I-1.4 :** histogrammes de distribution (effectifs en ordonnées, modalités en abscisses) des caractères de couverture des glumes (COV) ; texture de l'albumen (COR) ; date de floraison (FLK) et hauteur de plante (PHTK) et pour deux formes intermédiaires GC et DC, et les races principales à l'origine de ces formes hybrides.



intermédiaires amène à une remarque. L'hybridation entre les races concernées est possible et dans ces deux cas, une bonne recombinaison a été obtenue puisque les hybrides ont été maintenus et ont acquis une certaine importance.

La race intermédiaire GC a le plus fort index de diversité (0,72) (Annexe 1 et 2). Cette augmentation significative ( $p < 0,01$ ) de la diversité par rapport aux deux races de base - caudatum (0,67) et guinea (0,62) - peut résulter du fait que les recombinaisons se sont produites et les hybrides obtenus présentent la gamme de variabilité maximale permise (Figures I-1.4). Le croisement entre les races caudatum - grains très farineux dans 35% des cas - et guinea - grains très vitreux (19%) - a donné des formes intermédiaires au sein desquelles seul 1% des variétés sont à grains très vitreux et 13% à grains très farineux. Cette observation rend sans doute compte du fait que les hybrides obtenus par les programmes de sélection, visant à combiner la qualité des guinea avec les rendements élevés des caudatum, n'aient pas eu le succès escompté auprès des consommateurs et donc des paysans en Afrique occidentale. S'agissant des caractéristiques relatives au photopériodisme, la variance des caractères de date de floraison et de hauteur de la plante pour les variétés de la race GC est significativement supérieure à celle des races parentales ( $p < 0,01$ ) (test de Levene).

La race intermédiaire DC est distribuée essentiellement en Afrique orientale (64%). La diversité phénotypique de cette race (0,67) est du même ordre que celle de la race caudatum (Annexe 1 et 2). Par contre, elle est significativement supérieure à celle de la race durra (0,56). Les variétés précoces sont plus représentées dans la race intermédiaire que dans les deux races de base (39% contre 13 et 15%). Pour la race intermédiaire, la variance de la date de floraison est significativement supérieure à celle de chacune des races de base (test de Levene avec  $p < 0,01$ ). Par contre, pour la hauteur de la plante, la variance de la race DC n'est significativement supérieure que par rapport à la race durra. Les variétés à gros grains sont plus nombreuses à appartenir à la race DC (34%) qu'aux races caudatum (17%) ou durra (24%).

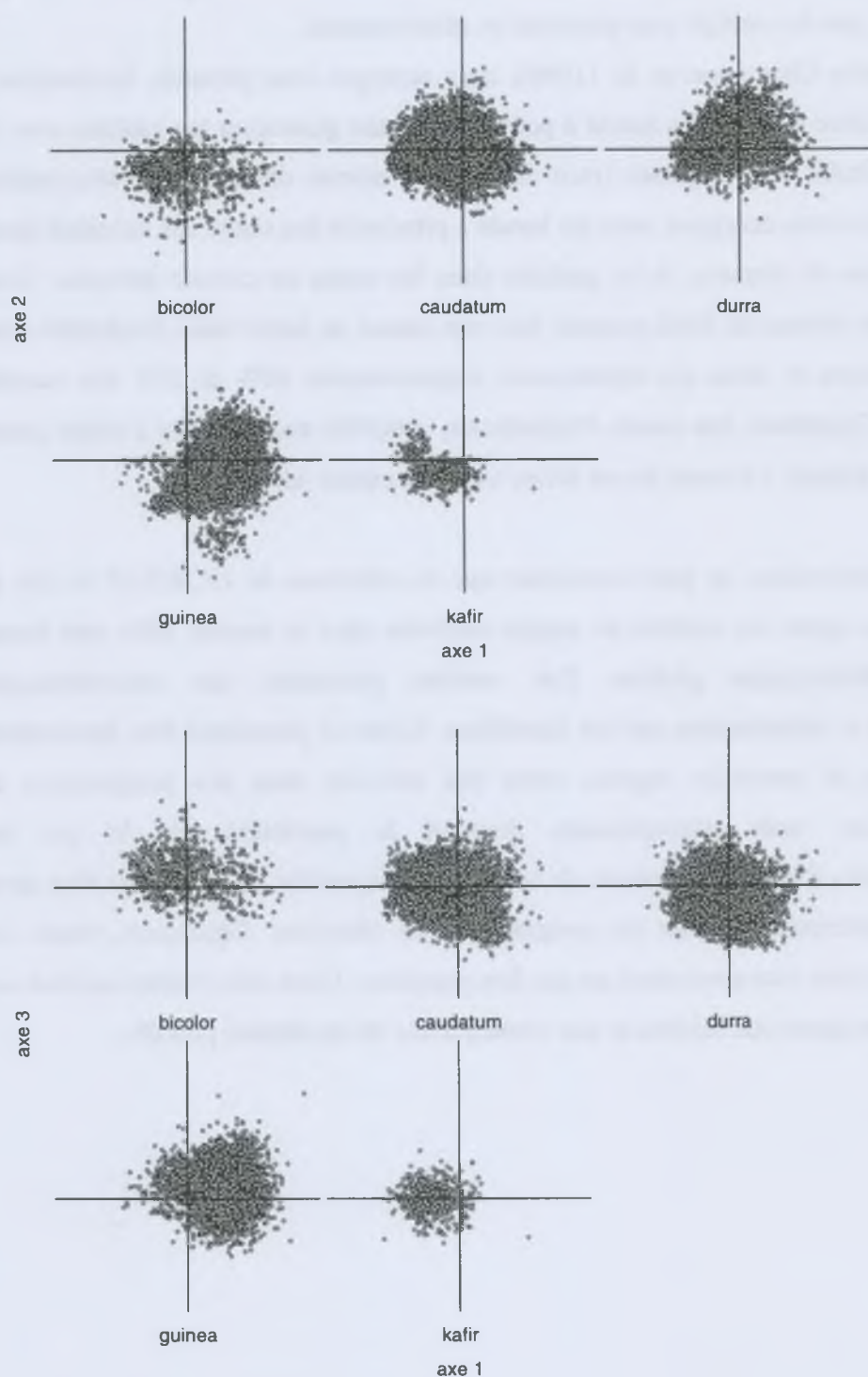
La situation des autres races intermédiaires est intéressante pour plusieurs raisons.

- leur faible représentation peut faire penser que ces hybrides ont peu de chances de se produire naturellement car les races majeures n'ont pas les mêmes exigences de développement. Par contre, celles-ci pourraient prendre de l'importance par le biais des programmes de sélection.
- étant donné la large distribution des races majeures, les hybridations se produisent mais, les hybrides obtenus de valeur inférieure à celle des parents sont contre sélectionnés.

- certaines combinaisons ne donnent pas des hybrides viables (système d'incompatibilité classique, incompatibilité structurelle des génomes, ...) par conséquent, peu de variétés sont produites et sélectionnées.

Au Mali, selon Chantereau et al. (1998), trois écotypes sont présents, les keninké à gros grains (race guinea), les kendé à petit grains (race guinea) et les gadiaba avec de grosses panicules et gros grains (race durra). Ces auteurs ont noté une structuration spatiale de ces trois écotypes, avec les kendé à proximité des cases, les kenenké plutôt dans les zones de brousse, et les gadiaba dans les zones de cuvette humides. Cette répartition au niveau du Mali pourrait être une raison du faible taux d'hybrides entre les races guinea et durra qui représentent respectivement 69% et 10% des variétés maliennes. Cependant, des essais d'hybridation contrôlée entre guinea à petits grains et durra ont échoué. La cause de cet échec n'est pas encore identifiée.

En conclusion, on peut considérer que la collection de l'ICRISAT si elle ne renferme pas toutes les variétés de sorgho cultivées dans le monde, offre une bonne diversité phénotypique globale. Des variétés présentant des caractéristiques intéressantes et inhabituelles ont été identifiées. Celles-ci pourraient être directement testées dans de nouvelles régions et/ou être utilisées dans des programmes de sélection. Les races intermédiaires donnent la possibilité par le jeu des recombinaisons d'élargir la gamme de variabilité disponible et pourraient être aussi des cibles intéressantes pour les programmes de sélection. Cependant, toutes les hybridations inter race pourraient ne pas être possibles. Cette observation soulève des interrogations quant aux raisons et aux conséquences de ces échecs putatifs.



**Figure I-2.1** : représentation dans l'espace bidimensionnel (plans 1x2 et 1x3) des différentes races de sorgho. Les axes sont déterminés par l'ACP réalisée sur les 22 473 accessions.

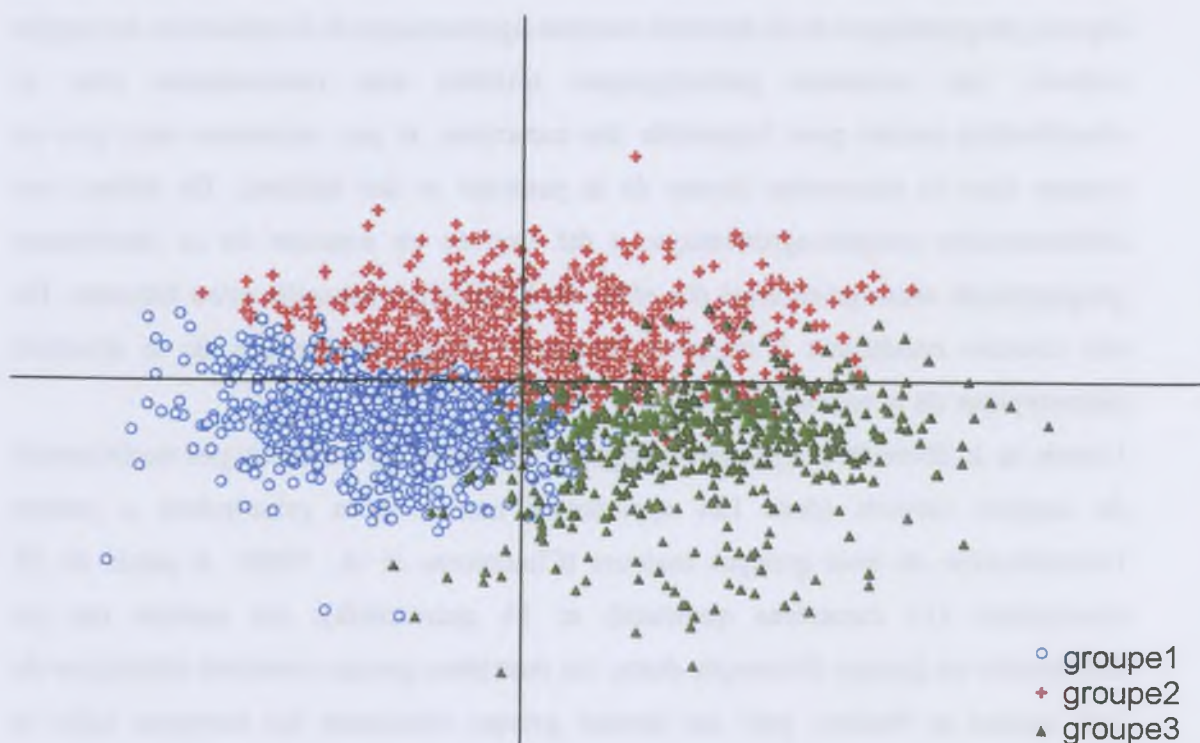


## CHAPITRE 2 : Organisation de la diversité

L' étude précédente a révélé une différenciation entre les races et entre les régions géographiques de la diversité morpho-agronomique de la collection de sorgho cultivés. Les tendances phénotypiques révélées sont concordantes avec la classification raciale pour l'ensemble des caractères, et pas seulement ceux pris en compte dans la taxonomie (forme de la panicule et des épillets). De même, une différenciation morpho-agronomique a été trouvée en fonction de la distribution géographique, témoignant ainsi des effets de la sélection naturelle et/ou humaine. De tels constats conduisent à suggérer l'existence d'une structuration de la diversité phénotypique de la collection.

L'étude de la diversité morphophysiolologique, effectuée sur 157 écotypes traditionnels de sorghos cultivés (dont 124 appartenant aux 5 races principales) a permis l'identification de trois groupes majeurs (Chantereau et al., 1989). A partir de 25 descripteurs (11 caractères qualitatifs et 14 quantitatifs), ces auteurs ont pu différencier un groupe d'écotypes durra, un deuxième groupe constitué d'écotypes de race guinea et bicolor, puis un dernier groupe réunissant les écotypes kafir et caudatum.

Dans ce chapitre, nous nous proposons d'étudier la structuration au sein des variétés traditionnelles de sorgho. Pour ce faire, une analyse en composantes principales (ACP) est réalisée sur l'ensemble des 22 473 accessions de la collection, en utilisant les 10 descripteurs morpho-agronomiques. Seuls sont retenus les facteurs qui ont une valeur propre supérieure à 1 (critère de Kaiser). Cette décision arbitraire vise à ne retenir un facteur que s'il est au moins équivalent à une variable originale. Les composantes principales sont extraites par une rotation qui maximise la variance sur les nouveaux axes (procédure "varimax" Statistica). Les trois premiers facteurs expliquent 61% de la variance morpho-agronomique. L'axe 1, est fortement corrélé aux caractères de date de floraison et hauteur de la plante. Le deuxième axe est influencé par les caractères de taille et poids des grains. Le troisième axe est associé à la taille de l'exertion pédonculaire. Seules les cinq races principales sont représentées dans l'espace bidimensionnel (Figure I-2.1). Cette représentation graphique des races prises séparément permet de différencier plus facilement leur tendance morpho-agronomique respective. L'axe 1, oppose nettement les guinea et les kafir. Les variétés guinea se caractérisent par des sorghos tardifs et de grande taille, alors que la race



**Figure I-2.2** : représentation de 2 000 accessions de sorgho classées dans trois groupes à l’aide des nuées dynamiques dans le plan 1x2 défini par l’ACP réalisée sur l’ensemble des 22 473 accessions.

**Tableau I-2.1** : répartition des 5 races principales dans les 3 groupes obtenus pas la méthode des nuées dynamiques.

	groupe 1	groupe 2	groupe 3	Total
bicolor	193	30	177	400
caudatum	171	180	49	400
durra	109	275	16	400
guinea	30	105	265	400
kafir	362	23	15	400
Total	865	613	522	2 000

kafir est surtout représentée par des variétés précoces et de petite taille. Pour les bicolor, caudatum et durra, les accessions semblent bien réparties dans le nuage de points. Le deuxième axe oppose les variétés à petits grains à celles à gros grains. Il permet de distinguer très clairement les races bicolor, durra et kafir des autres races. Les bicolor et kafir sont majoritairement caractérisés par de petits grains et les durra plutôt par de gros grains. Le troisième axe, associé à la taille de l'exertion pédonculaire, ne permet pas de discrimination très marquée. Les variétés des races bicolor et guinea ont tendance à être plus souvent caractérisées par un pédoncule légèrement plus long que la moyenne.

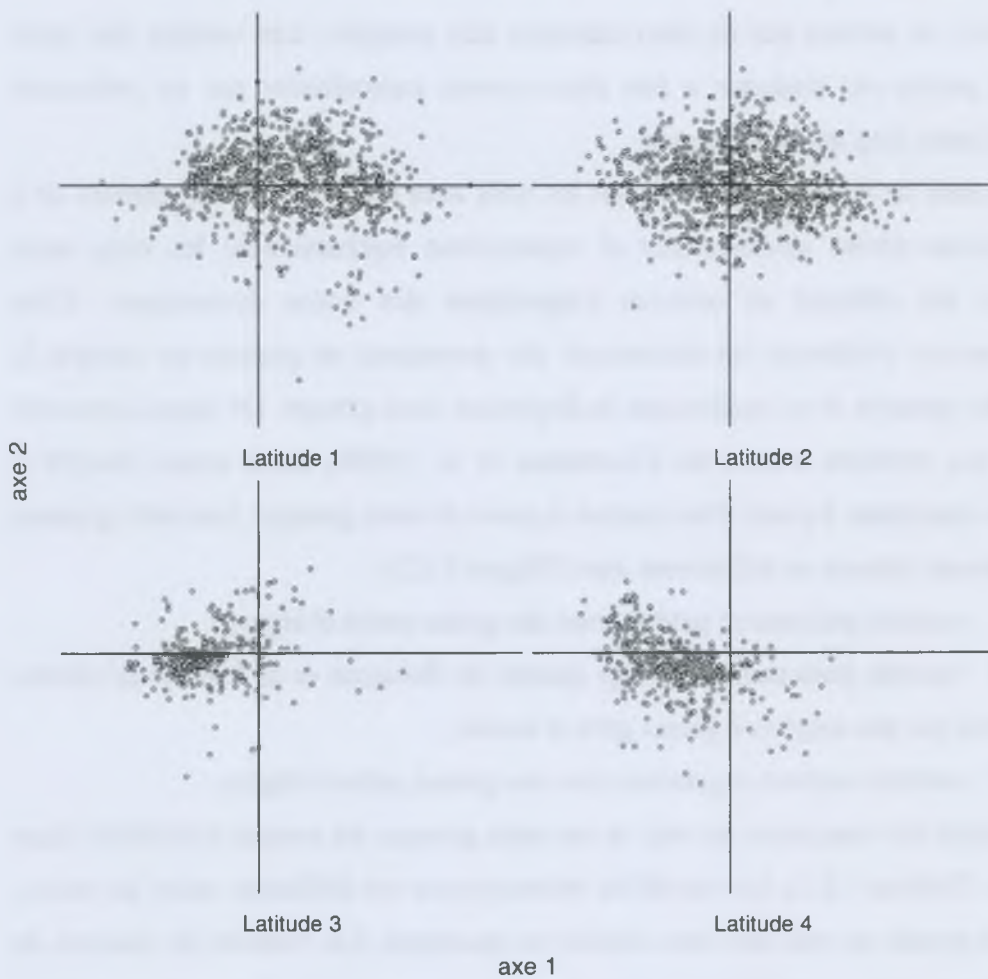
A partir de leurs coordonnées sur les trois axes de l'ACP, un classement de 2 000 accessions prises aléatoirement et représentant équitablement les cinq races principales, est effectué en utilisant l'algorithme des nuées dynamiques. Cette méthode permet d'effectuer un classement des accessions en prenant en compte la variance des groupes et en minimisant la dispersion intra groupe. De façon à pouvoir comparer nos résultats à ceux de Chantereau et al. (1989), nous avons cherché à obtenir un classement à partir d'un nombre à priori de trois groupes. Les trois groupes que nous avons obtenus se définissent ainsi (Figure I-2.2) :

- groupe 1 : variétés précoces et petites avec des grains petits et légers ;
- groupe 2 : variétés présentant une large gamme de floraison et de hauteur de plante, et caractérisé par des sorghos à grains gros et lourds ;
- groupe 3 : variétés tardives et grandes avec des grains petits et légers.

La distribution des cinq races au sein de ces trois groupes est ensuite considérée dans un tableau (Tableau I-2.1). La variabilité phénotypique est différente selon les races ; elle est très grande au sein des races bicolor et caudatum. Les variétés de chacune de ces deux races sont essentiellement classées dans deux groupes. Les variétés bicolor se répartissent bien dans le groupe 1 (48%) et dans le groupe 3 (44%), et les variétés de la race caudatum sont distribuées de manière équivalente dans le groupe 1 (43%) et dans le groupe 2 (45%).

A l'opposé, la variabilité est faible pour la race kafir dont les variétés appartiennent principalement au groupe 1 (90%). La race kafir comprend la majorité des sorghos précoces de petite taille et à petits grains (42%). Une situation intermédiaire est observée pour les races durra et guinea dont les 2/3 des variétés se situent dans un groupe : 69% dans le groupe 2 et 27% dans le groupe 1 pour la race durra, et 66% dans le groupe 3 et 26% dans le groupe 2 pour la race guinea. La race durra contribue





**Figure I-2.3 :** représentation de 2 000 accessions groupées par classe latitudinale sur le plan 1x2 défini par l'ACP réalisée sur les 22 473 accessions de sorgho.

largement aux variétés caractérisées par des grains gros (45%), alors que la race guinea inclut la majorité des plantes tardives, grandes et à petits grains (51%).

La diversité phénotypique de la collection montre une structuration des sorghos cultivés qui n'est pas en désaccord avec celle proposée par Chanterreau et al., (1989). En effet, les races durra, guinea et kafir sont bien différenciées les unes des autres, alors que bicolor et caudatum semblent plus diversifiées.

Avec seulement 10 caractères quantitatifs, (dont cinq sont communs avec ceux utilisés par Chanterreau et al.) et conformément aux résultats obtenus sur les 157 écotypes, nous pouvons confirmer une organisation de la collection des sorghos cultivés et corroborer ainsi les différences raciales au niveau des épillets avec celles observables au niveau de différences morphophysologiques globales. Cette relation résulte très probablement de l'action de sélection portant sur les associations entre caractères, et ayant structuré les formes raciales.

Dans les plans factoriels définis précédemment, la représentation des 2 000 accessions groupées en fonction de leur distribution latitudinale montre également quelques tendances générales (Figure I-2.3). Les sorghos équatoriaux auraient un cycle végétatif plutôt long (variétés tardives, et de taille élevée), avec une légère dominance pour des variétés à grains gros. Les sorghos cultivés dans les zones tropicales semblent bien répartis sur les axes corrélés au développement végétatif et aux caractéristiques des grains. En s'élevant dans les latitudes tempérées, les variétés tendent à devenir précoces et de petite taille, et les grains sont plus petits. Aucune différenciation géographique majeure n'est perceptible au niveau du 3<sup>ème</sup> axe. La distribution géographique des 2 000 accessions de sorghos montre des effectifs déséquilibrés entre les quatre classes latitudinales (34% latitude 1, 37% latitude 2, 13% latitude 3 et 16% latitude 4). Par un test de Levene prenant en compte le groupement géographique, la variance des caractères morpho-agronomiques entre les classes latitudinales est significativement différente ( $p < 0,05$ ).

Afin de tester l'existence de cette structuration raciale, nous avons prélevé aléatoirement un même effectif de 500 accessions pour chacune des races. Pour les 10 caractères quantitatifs accessibles dans la base de données, une ANOVA nous a permis de montrer que la variance inter-race est supérieure à la variance intra-race. Pour chaque caractère, excepté la hauteur de plante en saison rabi, le test de Levene révèle des différences de variance inter-race significatives ( $p < 0,05$ ). Pour le caractère

hauteur de la plante, seules les races caudatum et kafir ont des moyennes qui ne diffèrent pas significativement ( $p > 0,14$  par le test de Newman et Keuls). Une structuration raciale de la diversité morphométrique semble bien se confirmer par ces analyses de variances.

Pour les caractères qualitatifs, la distribution des 2 500 accessions de sorghos dans les différentes modalités est comparée par des tests de chi-2 en groupant les variétés par race, ou par latitude ou par race et latitude. Les distributions observées sont significativement différentes de celles attendues dans l'hypothèse d'une indépendance entre la classification raciale, la distribution latitudinale et les caractères morpho-agronomiques.

L'ensemble de ces comparaisons - analyses de variances pour les caractères quantitatifs ou tests de chi-2 pour les caractères qualitatifs - a montré clairement que les caractéristiques morpho-agronomiques des variétés de sorghos ne sont indépendantes ni de la classification taxonomique ni de la distribution latitudinale. Ainsi une organisation de la diversité phénotypique de la collection des sorghos cultivés peut être envisagée par la considération simultanée des races et des quatre classes de latitude.

Au cours de cette première partie sur la diversité de la collection des sorghos maintenus à l'IAC, nous avons pu montrer que les informations contenues dans la base de données sont exploitables pour l'étude de la diversité des sorghos cultivés. Aussi, une importante diversité phénotypique a été trouvée au sein des variétés traditionnelles de sorgho. Leur distribution couvre une large zone géographique mais reste tout de même légèrement sous-représentée pour certaines régions, comme l'Afrique australe et certains pays, comme la Chine et quelques pays d'Afrique. Les races sont également plutôt bien représentées bien que quelques unes contribuent moins à la collection. Les tendances phénotypiques majeures des races, s'accordent avec la classification raciale. Toutefois quelques formes originales ont pu être mises en évidence. Par leur importance au sein de la collection, quelques races intermédiaires ont fait preuve de succès au près des utilisateurs (consommateurs et agriculteurs). Cependant, certaines combinaisons, encore rarement représentées, posent des questions sur leur succès et/ou intérêts.

La diversité de la collection est apparue organisée en fonction de la classification raciale non seulement par rapport aux caractères utilisés dans cette classification des sorghos (structure de la panicule et de l'epillet), mais également du



point de vue de la diversité phénotypique globale. Cette généralisation d'une structuration raciale pourrait résulter d'associations entre caractères qui sont soumis à la sélection (naturelle et/ou humaine). L'origine géographique est également un facteur structurant la diversité des sorghos cultivés par le biais des adaptations environnementales spécifiques, comme la distribution latitudinale et/ou des adaptations aux pressions de sélection humaine voire culturelle.

En conclusion, la collection de sorghos maintenue à l'IAC peut répondre à son statut de collection mondiale. Il s'agit maintenant de considérer l'accès et une plus large utilisation d'une telle collection (22 473 accessions). La deuxième partie de notre travail va donc se concentrer sur la façon de procéder pour faciliter la gestion et l'utilisation des ressources génétiques du sorgho.

**2<sup>ÈME</sup> PARTIE - CORE COLLECTIONS**

**2<sup>ÈME</sup> PARTIE - CORE COLLECTIONS**

Nous avons vu précédemment que différentes procédures d'échantillonnage peuvent être appliquées pour la constitution de core collections prenant en compte les objectifs assignés aux core collections et la nature des données disponibles concernant la collection de base.

Dans cette deuxième partie, nous nous proposons de tester différentes stratégies d'échantillonnage sur la collection des variétés traditionnelles de sorgho. Dans un premier temps, à partir de la collection de base, trois stratégies de prélèvement aléatoire ont été utilisées et nous avons évalué la représentativité des trois core collections ainsi constituées vis-à-vis de la diversité contenue dans la collection de base. Ces travaux ont fait l'objet de deux articles acceptés sous réserve de corrections dans *Crop Science*. Par la suite, une des trois core collections a été réduite à 1% et deux nouveaux sous-ensembles ont été constitués indépendamment à partir de stratégies de prélèvement sélectif. Ces trois sous-ensembles représentant environ 1% de la collection de base ont été comparés entre eux et à la collection de base en considérant la diversité morpho-agronomique. Les résultats obtenus ont conduit à la rédaction d'un article accepté pour publication dans *Theoretical and Applied Genetics*.



Tableau II-1.1 : détermination des quatre catégories de réaction à la photopériode

écarts	catégorie	FL = FLK - FLR	PHT = PHTK - PHTR
[min - Q1]	1	[-58 ; 4]	[-170 ; 65]
[Q1 - Médiane]	2	[4 ; 23]	[65 ;130]
[Médiane - Q3]	3	[23 ; 43]	[130 ; 180]
[Q3 - max]	4	[43 ; 139]	[180 ; 425]

FLK et PHTK = date de floraison et hauteur de plantes à la saison kharif (jours longs)  
FLR et PHTR = date de floraison et hauteur de plantes à la saison rabi (jours courts)  
min et max : valeurs minimale et maximale dans la collection de base  
Q1 et Q3 : premier et dernier quartiles

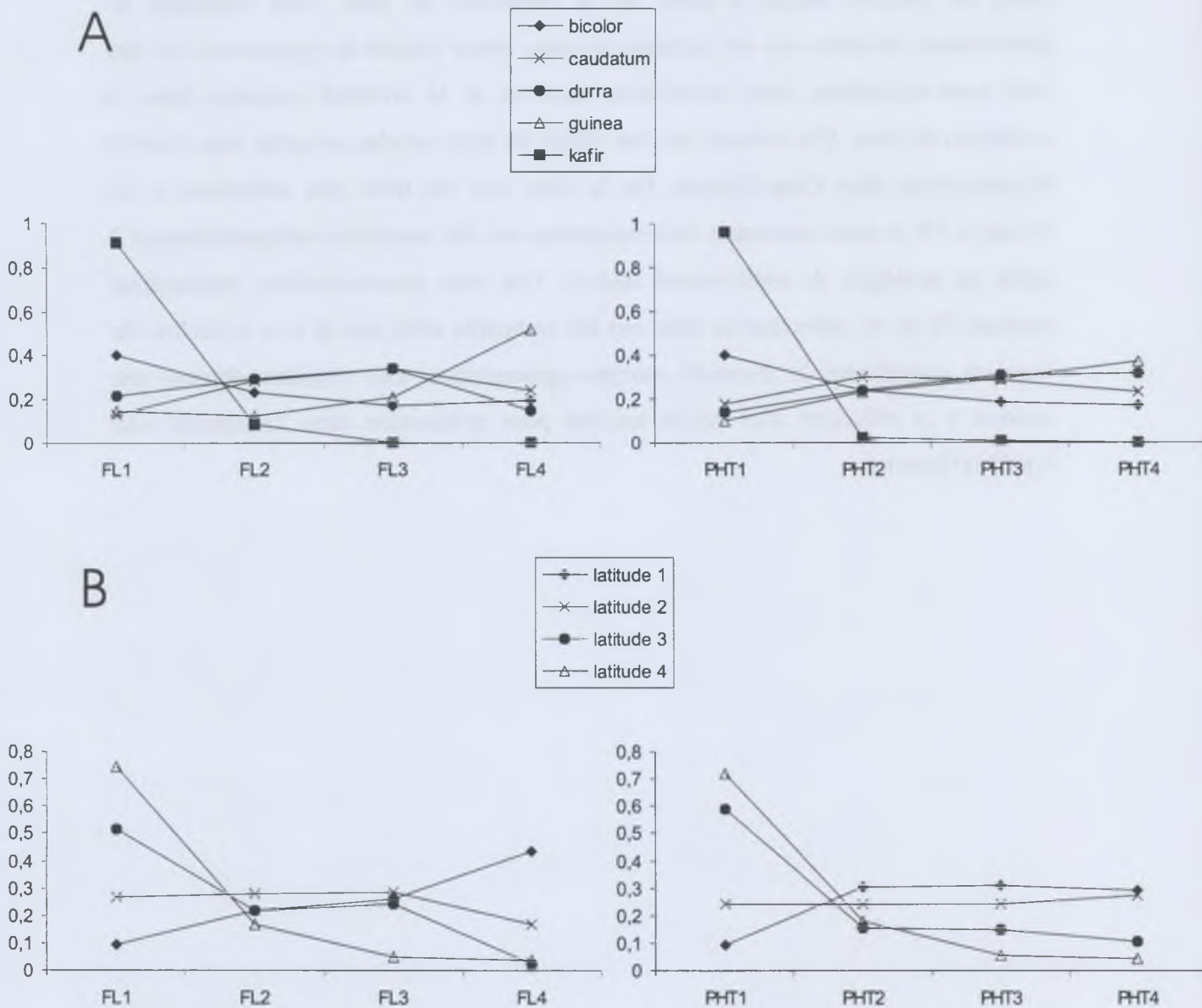


Figure II-1.1 : distribution des fréquences dans les quatre catégories de photopériodisme (définies dans le Tableau II-1.1) pour les 5 races de base (A) et les 4 classes latitudinales (B).

## CHAPITRE 1 : Stratégies de prélèvement aléatoire

Proposées par Brown (1989b), les stratégies d'échantillonnage aléatoire sont applicables sur des collections structurées en groupes non-chevauchants. Une première étape dans la réalisation de telles core collections est donc basée sur l'étude de l'organisation de la diversité de la collection et la mise en évidence d'une structuration. A partir des groupes ainsi obtenus, des prélèvements aléatoires sont effectués indépendamment ou non de la taille des groupes.

### I - Structure de la collection de base

Dans la première partie, l'analyse de la diversité morpho-agronomique de la collection nous a permis de mettre en évidence une organisation raciale de la diversité phénotypique. En considérant uniquement les races majeures et un nombre équivalent d'accessions par race (500), l'analyse factorielle résumée dans un espace bidimensionnel, a conduit à un premier facteur expliquant 30% de la variabilité totale et fortement corrélé aux caractères du cycle végétatif, date de floraison et hauteur de plante.

Les écarts dans le cycle végétatif, entre la saison des jours longs (kharif) et la saison des jours courts (rabi) témoignent de la sensibilité des variétés à la photopériode. Plus les écarts sont importants, plus les variétés sont photosensibles. A l'inverse, une variété photoinsensible est caractérisée par de faibles valeurs. Pour les caractères affectés par la photopériode - date de floraison et hauteur de plante - une nouvelle variable est obtenue à partir de la différence des valeurs enregistrées aux deux saisons, respectivement FL pour la date de floraison et PHT pour la hauteur de plante. Ces deux variables FL et PHT ont été découpées en quatre classes d'effectif à peu près égal (Tableau II-1.1). Ces quatre classes pourraient aussi correspondre aux quatre gènes majeurs impliqués dans le déterminisme floral, ou aux quatre gènes majeurs contrôlant la hauteur de plante (Doggett, 1988).

La distribution des races et des groupes de latitude dans les quatre classes phénotypiques de date de floraison (FL1 à FL4) et de hauteur de plante (PHT1 à PHT4) montre que (Figure II-1-1) :

- les résultats sont similaires pour les deux caractères considérés ;

**Tableau II-1.2** : distribution des races de sorgho dans les quatre catégories de photopériodisme.

	photo-insensible	moyennement photosensible	photosensible	très photosensible	Total
bicolor	0	252	260	206	718
caudatum	175	171	1263	3284	4893
durra	0	0	4545	0	4545
guinea	0	0	148	3759	3907
kafir	514	6	0	0	520
CB	200	0	560	0	760
DB	0	31	599	632	1262
GB	0	4	10	206	220
KB	36	15	4	0	55
DC	0	189	1853	671	2713
GC	0	361	1200	838	2399
KC	142	9	53	0	204
GD	0	21	92	25	138
KD	69	0	43	0	112
GK	24	3	0	0	27
Total	1160	1062	10630	9621	22473



- les races bicolor, caudatum et durra sont relativement bien distribuées dans les quatre catégories phénotypiques ;
- de façon claire, les kafir se distinguent des autres races par leur présence quasi exclusive dans la catégorie des variétés photo-insensibles (FL1 et PHT1) ;
- à l'inverse, la race guinea est davantage présente dans les catégories de variétés photosensibles (FL4 et PHT4) ;
- les variétés originaires de la latitude 1 (équatoriale), sont rarement très petites et présentent en majorité de grands écarts quant aux dates de floraison aux deux saisons ;
- à l'inverse, aux latitudes tempérées (3 et 4), les variétés sont plutôt petites et photo-insensibles.

Une Classification Ascendante Hiérarchique (CAH) a été obtenue en calculant la distance Euclidienne entre les différents groupes race-latitude (60 groupes) à partir du tableau de contingence des distributions de fréquences de chaque groupe race-latitude dans les quatre classes phénotypiques pour FL et PHT. La procédure d'agrégation utilise la méthode de Ward, avec pour objectif de minimiser la somme des carrés de tous les couples (hypothétiques) de classes pouvant être formés à chaque étape. Le dendrogramme obtenu indique une structuration des 60 groupes en quatre classes de photopériodisme. A partir de cet a priori de quatre classes, l'utilisation des nuées dynamiques (minimisation des variances intra-classe et maximisation de la variance inter-classe), conduit à un classement des 60 groupes à 95% en accord avec celui obtenu par la CAH. Ce résultat suggère une bonne robustesse de la structuration des sorghos selon le critère de photosensibilité. Les quatre classes obtenues par les nuées dynamiques ont des effectifs très inégaux (Tableau II-1.2). La distribution des 15 races dans les quatre classes révèle que aucun kafir n'est photosensible, tous les durra sont photosensibles et les guinea à plus de 95% très photosensibles. Cette structuration des sorghos traduit l'adaptation éco-géographique des variétés dans leur région d'origine, et l'adaptation aux pressions de sélection imposées par les agriculteurs. En effet, ces derniers sélectionnent fortement les variétés qui atteignent leur stade de maturation à la période la plus favorable pour éviter les moisissures ou les attaques de ravageurs (insectes et oiseaux).

Cette hiérarchisation des sorgho cultivés est intéressante à plusieurs points de vue :

- elle intègre à la fois une composante raciale (la pertinence de la classification raciale a été démontrée par plusieurs auteurs (Chanterreau et al., 1989 ; Deu et al., 1994)) et une composante géographique ;
- elle peut être utile pour les programmes de sélection. En effet, les sélectionneurs cherchent souvent à allier la souplesse d'adaptation climatique des variétés photosensibles, aux bonnes aptitudes productives (meilleur rapport grain/paille) des variétés photo-insensibles (Trouche et al., 1998) ;
- le photopériodisme, associé aux adaptations éco-géographiques et aux pressions sélectives humaines, permet de mettre en évidence une organisation spatiale de la diversité génétique de la collection. Une telle structuration est suggérée par Sokal, (1986, cité dans Balfourier et Charmet, 1994). Par la suite, l'échantillonnage en vue de la constitution de core collections est effectué à partir des quatre groupes de réponse à la photosensibilité.

## **II - Comparaison entre trois stratégies d'échantillonnage aléatoire**

Trois stratégies d'échantillonnage aléatoire (prélèvement constant C, proportionnel P et logarithmique L) ont été appliquées pour obtenir des core collections retenant 10% des accessions de la collection de base.

La comparaison des trois core collections (C, P et L) à la collection de base montre que la diversité morpho-agronomique globale caractérisée par l'indice de Shannon-Weaver et retenue dans ces core collections est forte ( $SDI = 0,835 \pm 0,050$  pour la core C,  $0,851 \pm 0,052$  pour la core P et  $0,840 \pm 0,050$  pour la core L) et comparable à celle de la collection de base ( $SDI = 0,853 \pm 0,053$ ). De plus, cette diversité ne diffère pas significativement entre les trois core collections ( $p < 0,05$ ). En considérant la distribution de chaque caractère, on note que la core collection P représente de façon fidèle la collection de base, tant pour les caractères morpho-agronomiques qualitatifs que pour les caractères quantitatifs. De même, la représentation des races comme celle des origines géographiques n'est pas significativement différente de celles dans la collection de base ( $p < 0,05$ ).

A l'inverse, la core collection C présente une distribution différente pour les caractères morpho-agronomiques, races et origines géographiques. Par construction, cette stratégie C échantillonne davantage en proportion dans les plus petits groupes ce qui

**Tableau II-1.3** : distribution des accessions au sein de chacune des trois core collections et dans chaque classe de photopériodisme en fonction de leur appartenance raciale (A) et de la latitude du pays d'origine (B).

I : insensible - MS : moyennement sensible - S : sensible - TS : très sensible

A	core collection C					core collection P					core collection L				
	I	MS	S	TS	Total	I	MS	S	TS	Total	I	MS	S	TS	Total
bicolor	0	140	17	12	169	0	33	25	14	72	0	117	16	12	145
caudatum	91	95	73	196	455	26	10	113	319	468	71	75	79	201	426
durra	0	0	237	0	237	0	0	461	0	461	0	0	263	0	263
guinea	0	0	11	208	219	0	0	23	391	414	0	0	3	266	269
kafir	241	1	0	0	242	42	0	0	0	42	220	1	0	0	221
CB	95	0	37	0	132	22	0	65	0	87	90	0	38	0	128
DB	0	16	29	40	85	0	3	64	53	120	0	15	38	36	89
GB	0	2	0	12	14	0	0	2	17	19	0	1	1	14	16
KB	22	7	0	0	29	3	2	0	0	5	13	5	0	0	18
DC	0	78	98	35	211	0	15	180	71	266	0	92	113	34	239
GC	0	203	52	57	312	0	39	119	92	250	0	165	78	70	313
KC	69	5	2	0	76	11	1	2	0	14	60	2	4	0	66
GD	0	14	5	2	21	0	3	8	5	16	0	7	8	2	17
KD	32	0	1	0	33	11	0	1	0	12	23	0	1	0	24
GK	12	1	0	0	13	1	0	0	0	1	11	2	0	0	13
Totals	562	562	562	562	2248	116	106	1063	962	2247	488	482	642	635	2247

B	core collection C					core collection P					core collection L				
	I	MS	S	TS	Total	I	MS	S	TS	Total	I	MS	S	TS	Total
latitude 1	1	89	437	6	533	0	154	732	1	887	1	80	485	3	569
latitude 2	27	423	125	7	582	6	806	230	2	1044	30	515	150	5	700
latitude 3	322	45	0	252	619	68	81	0	44	193	262	40	0	212	514
latitude 4	212	5	0	297	514	42	22	0	59	123	195	7	0	262	464
Total	562	562	562	562	2248	116	1063	962	106	2247	488	642	635	482	2247



conduit à déséquilibrer fortement la représentation des différents groupes lorsque ceux-ci sont d'effectifs très inégaux (Tableau II-1.3). Dans notre étude, la fréquence des sorghos photo-insensibles et moyennement photosensibles est passée de 5% dans la collection de base à 25% dans la core collection C. De tels changements dans les fréquences de distribution ont conduit à une modification majeure des distributions des caractères morpho-agronomiques. Ainsi, tous les caractères qualitatifs sont significativement affectés par l'échantillonnage ( $p < 0,05$ ) excepté la pigmentation foliaire qui est une variable très peu polymorphe (une de ses deux modalités représente 98% de la collection de sorgho). La variance est significativement diminuée pour tous les caractères quantitatifs, sauf pour la date de floraison à la saison kharif et la largeur de la panicule. Pour cette dernière, la moyenne dans la core collection est significativement augmentée ( $p < 0,05$ ), alors que celle pour la date de floraison est significativement diminuée dans la core C. Cette dernière observation traduit le biais dans la représentation en faveur des formes photo-insensibles. Par rapport à la collection de base, toutes les races exceptées les races *caudatum*, GB et GD, ont une distribution significativement différente. Ces trois races présentent une bonne gamme de photopériodisme ; ceci peut expliquer le faible impact de l'échantillonnage constant sur leur distribution. La représentation des trois continents est significativement différente dans la core collection C. Les sorghos originaires d'Afrique et d'Asie sont en proportion plus faible que dans la collection de base et ceux du groupe "Amérique-Océanie-Europe" sont sur-représentés. Ceci s'explique en grande partie par la distribution latitudinale des sorghos pour ces trois continents. En effet, la distribution des quatre classes de latitude est modifiée, avec une sous-représentation des zones équatoriales et tropicales, là où les sorghos sont essentiellement photosensibles.

La core L se situe à une position intermédiaire entre la core P et la core C. Elle se caractérise par une fréquence plus élevée en variétés photo-insensibles ou moyennement photosensibles que dans la collection de base tout en ayant une bonne représentation des autres groupes de photopériodisme (Tableau II-1.3). La core L diffère significativement de la collection de base pour tous les caractères morpho-agronomiques qualitatifs exceptées la pigmentation foliaire (dû au manque de polymorphisme du caractère en question) et la lustrosité du grain. La variance est significativement diminuée pour tous les caractères quantitatifs exception faite pour la hauteur de plante en kharif où elle est augmentée, et pour la date de floraison en

kharif et la largeur de la panicule où la variance n'est pas significativement modifiée. Cette homogénéité des variances nous permet de comparer les moyennes, et de montrer ainsi une diminution significative ( $p < 0,05$ ) de la moyenne de la date de floraison et une augmentation significative de la largeur moyenne de la panicule. Concernant la représentation raciale, la core L diffère significativement de la collection de base pour toutes les races sauf GB, DC et GD. Ces intermédiaires sont caractérisées par une assez bonne gamme de photopériodisme ce qui pourrait expliquer l'absence d'effet de l'échantillonnage logarithmique compensant la sur-représentation de leur formes peu ou pas photosensibles par une sous-représentation de leur formes photosensibles et très photosensibles. Pour la distribution géographique des accessions prélevées dans la core L, seule la représentation des sorghos asiatiques n'est pas significativement modifiée par rapport à la collection de base. Les variétés équatoriales et tropicales sont significativement diminuées, et les variétés tempérées augmentées. Par rapport à la collection de base, et comme dans l'échantillonnage précédent, la stratégie L favorise la représentation des formes photo-insensibles. Cependant, la core L diffère de la core C par une plus grande exploration des caractères associés au photopériodisme. La variance de la date de floraison, ainsi que les moyennes pour la hauteur de plante en kharif et rabi, et pour le poids des grains sont significativement supérieures dans la core L. De plus, la stratégie L représente significativement plus de variétés tropicales et moins de variétés tempérées que la core C. Aussi, un nombre significativement plus grand de variétés guinea est retenu dans la core L. Ces différences qui viennent d'être soulignées révèlent un échantillonnage moindre des variétés photosensibles et très photosensibles dans le cas de la stratégie L.

En conclusion, la stratégie P conduit à une core collection image réduite de la collection de base. Les stratégies C et L conduisent à une représentation de la diversité différent de celle de la collection de base. De tels résultats ont été également notés par Brown (1989b) pour les core collections de soja, Spagnoletti Zeuli et Qualset (1993) pour les core collections de blé dur, Diwan et al. (1995) pour les core collections de luzerne annuelle des USA, et Galwey (1995) pour les core collections de haricot. Dans notre étude, les trois stratégies conduisent à des core collections qui se différencient les unes des autres principalement par la distribution des caractères phénotypiques associés aux facteurs pris en compte lors de la stratification de la collection de base.



**Encadré 2 : la stratégie PCS ou Principal Component Scoring (Noirot et al., 1996 et Hamon et al., 1998)**

Cette stratégie a été proposée afin de maximiser la diversité retenue au sein de la core collection. Pour un échantillonnage optimal des individus les moins redondants, la stratégie PCS requiert deux conditions :

- 1 - dans l'ensemble échantillonné, tous les individus sont interféconds et deux phénotypes extrêmes peuvent reproduire la gamme des phénotypes intermédiaires
- 2 - la méthode est d'autant plus efficace que les caractères quantitatifs sont additifs et héréditaires. Dans la pratique, même une hérédibilité faible justifie l'utilisation de la méthode.

Dans une population de taille **N**, évaluée pour **J** descripteurs, la diversité peut être déterminée par la différence existant entre les individus et estimée par une distance.

La stratégie se déroule en plusieurs étapes:

- 1. une analyse multivariée permet d'éliminer les effets de colinéarité entre les caractères corrélés, et donne **K** nouvelles variables centrées et statistiquement indépendantes : les facteurs. Les données initiales sont ainsi transformées en coordonnées factorielles.
- 2. un calcul des distances entre les individus à partir des coordonnées factorielles. Ces dernières sont pondérées par les racines carrées des valeurs propres des facteurs correspondants, donnant ainsi le même poids à tous les facteurs dans l'estimation de la distance.
- 3. la définition des contributions des individus à la somme des carrés généralisée (**SSG**) obtenue par le produit **N x K** (Lebart et al., 1977 cité dans Noirot et al., 1996).

La contibution **P<sub>i</sub>** d'un individu **i** à la **SSG** correspond à la somme des carrés des coordonnées sur les **K** nouveaux axes :

$$P_i = \sum_{j=1}^K x_{ik}^2$$

On définit alors la contribution relative **CR<sub>i</sub>** de chaque individu **i** à la **SSG** de l'ensemble des données :

$$CR_i = P_i / (N \times K)$$

Le choix des individus se fait en fonction de leur contribution relative à la **SSG**. Le premier individu sélectionné est celui présentant la plus forte **CR<sub>i</sub>**. La procédure de sélection itérative des individus maximisant la diversité de la core collection est achevée lorsque la taille de la core collection et/ou le pourcentage de diversité retenue ont atteint les objectifs fixés.

Initialement développée pour les variables quantiatives, cette stratégie est également applicable aux caractères qualitatifs. Seules les deux premières étapes sont différentes :

	Caractères quantitatifs	Caractères qualitatifs
1ère étape analyse multivariée	Analyse en composantes principales	Analyse factorielle des correspondances
2ème étape calcul de distances	Distance Euclidienne	Distance du chi-2

Ces deux options de la stratégie d'échantillonnage PCS sont en accès libre à l'adresse suivante : <http://mpl.ird.fr/genetrop/rg/pcss/index.htm>



Finalement, compte tenu des caractéristiques des trois core collections, la core L ayant une représentation plus équilibrée des groupes de photosensibilité tout en respectant l'inégalité d'effectif de ces groupes dans la collection de base, nous a semblé plus particulièrement intéressante pour la poursuite de notre travail.

## ***CHAPITRE 2 : Impact des stratégies d'échantillonnage aléatoire vs. non aléatoire sur la diversité morpho-agronomique***

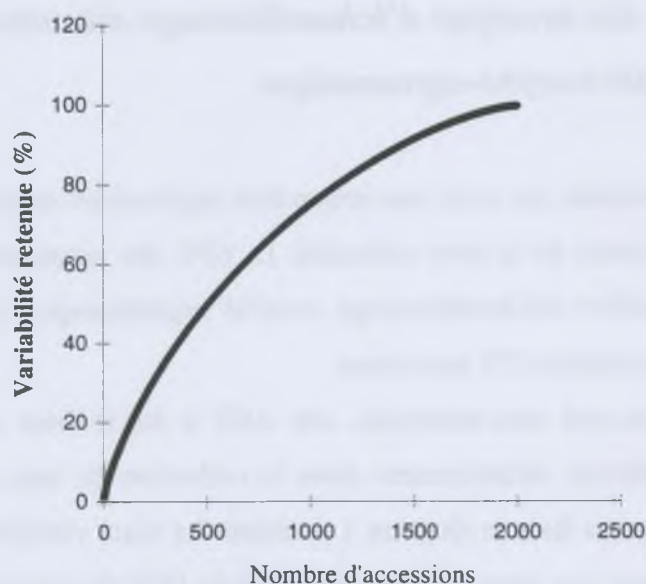
Cette étude a été réalisée sur trois sous-ensembles représentant environ 1% de la collection de base. A partir de la core collection L, 10% des accessions ont été prélevées suivant la procédure d'échantillonnage stratifié logarithmique, constituant ainsi le sous-ensemble L comptant 225 accessions.

Pour la constitution du second sous-ensemble, une ACP a été réalisée sur 2 000 accessions de sorgho prélevées aléatoirement dans la collection de base et les 12 variables quantitatives décrites dans le chapitre 1 (incluant les deux variables date de floraison et hauteur de plante aux deux saisons). La méthode PCS (Encadré 2) choisie pour maximiser la représentation de la diversité de la collection des 2 000 accessions dans un sous-ensemble de taille à définir, a permis de sélectionner 229 accessions (sous-ensemble PCS) contribuant à 31% de la variabilité totale (Figure II-2.1).

Le troisième sous-ensemble (T) a été constitué en considérant des critères taxonomiques et écogéographiques et des utilisations spécifiques des variétés. Ce sous-ensemble comprend 246 accessions.

L'impact des stratégies d'échantillonnage aléatoires vs. non aléatoires sur la diversité agromorphologique a été ensuite étudié.

La diversité phénotypique globale de la collection de base et pour chacun des sous-ensembles L, PCS et T est estimée à partir des analyses multivariées et des indices de diversité de Shannon-Weaver. D'un point de vue général, la diversité phénotypique est élevée et elle n'est pas significativement modifiée quelle que soit la stratégie utilisée (respectivement  $SDI = 0,853 \pm 0,053$ ,  $0,823 \pm 0,052$ ,  $0,853 \pm 0,048$  et  $0,826 \pm 0,051$  pour la collection de base, et les sous-ensembles L, PCS et T). Cependant, considérant individuellement chaque caractère morpho-agronomique et les informations de classification raciale et d'origine géographique, des différences



**Figure II-2.1 :** contribution des accessions à la variabilité phénotypique

vis-à-vis de la collection de base et entre les trois sous-ensembles sont mises en évidence.

De façon générale, le sous-ensemble PCS représente très fidèlement la collection de base, tant du point de vue des caractères morpho-agronomiques qualitatifs que pour les distributions raciales et géographiques. Ce fait souligne clairement que les caractères non pris en compte dans la mise en oeuvre de cette stratégie ne sont pas affectés par l'échantillonnage. Quelques variables quantitatives présentent des propriétés statistiques modifiées. Ce sont celles corrélées aux deux premiers facteurs de l'ACP, modifiées soit par une augmentation de leur variance (pour la hauteur de plante en saison rabi et le poids de 100 grains), soit par une diminution de leur moyenne (dans le cas de la taille des grains). Les autres variables quantitatives ne sont pas affectées autrement que par une diminution du coefficient de Kurtosis c'est-à-dire un aplatissement de leur distribution.

La représentation géographique du sous-ensemble PCS n'est pas significativement différente de celle de la collection de base. Cependant, pour la distribution raciale, les modifications les plus importantes sont notées pour certaines races intermédiaires (non représentées dans le sous-ensemble PCS). Ceci est à considérer avec l'hypothèse génétique de départ formulée dans la stratégie PCS. A savoir, au sein du groupe à échantillonner, il n'y a aucune barrière génétique à la reproduction et toutes les combinaisons génétiques peuvent être obtenues à partir d'un minimum d'individus.

Des observations similaires ont également été faites lors de l'étude de quatre plantes (sorgho, riz, hévéa et caféier) par Hamon et al. (1998). Ces auteurs ont conclu à la non influence de la stratégie PCS sur les caractères non pris en compte dans la mise en oeuvre de la stratégie. Dans la plupart des core collections établies, la variance pour une majorité de caractères a été significativement augmentée. Une exception a toutefois été relevée dans le cas des caféiers, où relativement peu de variances ont été affectées mais où l'effet principal de la PCS s'est caractérisé par un aplatissement de la distribution.

A l'opposé du sous-ensemble PCS, les deux autres sous-ensembles s'écartent de la distribution observée dans la collection de base à la fois pour les caractères morpho-agronomiques et pour les informations raciale et géographique.

Le sous-ensemble L diffère de la collection de base pour des caractères morpho-agronomiques associés au photopériodisme (cycle végétatif pendant la saison kharif). Les caractères largement affectés par les conditions environnementales



(nombre de talles racinaires), ou résultant d'adaptations géographiques (largeur et forme de la panicule, recouvrement des glumes), ou encore ceux soumis aux pressions humaines pour le rendement ou la qualité des grains (taille et poids des grains, couleur des grains, texture de l'albumen) sont aussi modifiés. Il en est de même pour la distribution des races et des origines géographiques caractéristiques des variétés photo-insensibles ou moyennement photosensibles. Leur représentation dans le sous-ensemble L est significativement plus élevée que dans la collection de base (latitudes tempérées, races bicolor et kafir). Cependant, il peut paraître surprenant de voir que dans ce sous-ensemble la fréquence de la race GC ait été augmentée. En effet, les variétés GC sont essentiellement photosensibles et très photosensibles (respectivement 50% et 35% de la collection de base). Le sous-échantillon L a été réalisé à partir de la core collection L comprenant 2 247 accessions et dans laquelle la race GC était représentée à 53% par des variétés moyennement photosensibles. Ainsi, la seconde étape de l'échantillonnage, qui a permis d'obtenir 225 accessions, s'est effectué à partir d'une core collection non représentative de la collection de base concernant la distribution de la photosensibilité et où l'essentiel des variétés GC étaient peu photosensibles. Cet échantillonnage à deux étapes a conduit, *in fine*, à une bonne représentation des sorghos GC.

Il semble donc que les facteurs directement ou indirectement impliqués dans la stratification de la collection soient affectés par le prélèvement, de même que les caractéristiques qui leur sont associées.

La stratégie T établie sur les critères éco-géographiques et taxonomiques prend en compte un grand nombre de facteurs. Par conséquent, le sous-ensemble T diffère très largement de la collection de base pour de nombreux caractères morpho-agronomiques de même que pour les distributions raciales et latitudinales. Seules, les propriétés statistiques de la hauteur de plante et de la largeur de la panicule ne sont pas significativement modifiées. Autrement dit, l'échantillonnage représente relativement bien la gamme de variabilité de ces deux caractères. Pour les autres caractères, comme ceux associés au rendement (taille et poids des grains) ou au photopériodisme (date de floraison et hauteur de plante à la saison kharif), les variances sont significativement augmentées dans la core T. En ce qui concerne les caractères qualitatifs, dans tous les cas, la distribution est significativement modifiée. Les caractéristiques associées à l'adaptation à la sécheresse - panicule compacte (plus de 50% de la core T) - aux caractéristiques de qualité - glumes et/ou grains clairs

(respectivement 45% et 62% dans la core T), grains vitreux et/ou sans sous-couche brune (respectivement 78% et 82% dans la core T) - sont mieux représentées dans la core T que dans l'ensemble de la collection de base. Pour la représentation des origines latitudinales, seules les régions tropicales ont maintenu une fréquence comparable à celle de la collection de base. Les races qui sont sur-représentées dans la core T sont celles qui paraissent être essentielles dans l'alimentation humaine en Inde (durra) ou celles relativement peu sensibles ou insensibles au photopériodisme (bicolor, kafir et ses formes intermédiaires).

En conclusion, cette stratégie T semble être celle qui modifie le plus la distribution de la diversité génétique morpho-agronomique.

En comparant les trois stratégies entre elles, les tendances décrites ci-dessus sont confirmées. Les trois stratégies se différencient les unes des autres par des distributions significativement différentes pour le caractère de couverture des glumes (le sous-ensemble PCS majoritairement composée de variétés à grains peu ou pas recouverts, le L avec des variétés à grains moyennement à très recouverts, et le sous-ensemble T avec une très bonne représentation de toute la gamme de recouvrement) et la représentation des sorgho équatoriaux (respectivement 44%, 21% et 30% pour les sous-ensembles PCS, L et T).

Le sous-ensemble PCS se distingue des deux autres par une augmentation de la variance du caractère de hauteur de plante à la saison rabi et une sous-représentation de la race kafir. Ces caractéristiques s'accordent avec une représentation réduite des variétés photo-insensibles. La stratégie L diffère des stratégies PCS et T par une représentation moindre des variétés à panicule compacte, et une plus forte représentation des variétés à panicule lâche et branches érigées. De plus, ce sous-ensemble L se caractérise par des sorghos dont le cycle végétatif à la saison kharif est court (variétés en moyenne plus précoces et plus petites) et dans l'ensemble, la variance du caractère de poids de grains est significativement plus petite. Aussi, cette stratégie tend à sous-représenter la race durra et à sur-représenter les variétés tempérées (rang latitudinal 3). Ces observations corroborent bien le fait que la stratégie L favorise l'échantillonnage des variétés photo-insensibles. Le sous-ensemble T se démarque très largement des deux autres pour les caractères qualitatifs. Ainsi, cette stratégie T tend à orienter l'échantillonnage vers une meilleure représentation des variétés caractérisées par une nervure médiane de la feuille de

couleur verdâtre ou jaune. Ce sous-ensemble comporte davantage de sorghos à grains clairs et/ou lustrés et/ou sans sous-couche brune. La variance pour la longueur de la panicule et la taille des grains est plus grande dans ce sous-ensemble. Enfin, plus de variétés GC sont prélevées dans cette sélection. Ce sous-ensemble T diffère significativement des deux autres en de nombreux points, ce qui pourrait résulter par le fait qu'un grand nombre de critères sont pris en compte dans cette stratégie.

D'après ces résultats il paraît clairement que la diversité phénotypique globale n'est pas modifiée par les trois stratégies, mais que des différences dans la distribution des caractères morpho-agronomiques existent. Deux hypothèses peuvent alors être avancées ; ou ces différences sont le résultat d'un biais lors du prélèvement, ou bien il s'agit d'un effet dû à la mise en oeuvre de la stratégie d'échantillonnage utilisée. Dans le premier cas, la différence entre les trois sous-ensembles devrait encore s'exprimer en considérant d'autres jeux de caractères comme ceux liés aux résistances à des stress biotiques ou abiotiques ou même des marqueurs moléculaires. Dans la seconde hypothèse, si les modifications observées sont liées à la procédure d'échantillonnage de la collection de base, alors la diversité moléculaire, théoriquement indépendante des critères considérés lors de la constitution des trois sous-ensembles, ne devrait pas être ou peu modifiée.

Dans cette partie nous avons pu rendre compte des différences entre des stratégies d'échantillonnage aléatoire. La structuration de la collection de base semble être le facteur clé pour différencier les core collections. Quand une organisation de la diversité de la collection peut être mise en évidence, et lorsque les classes obtenues sont d'effectifs inégaux, les stratégies d'échantillonnage C et L permettent une bonne représentation de toutes les classes. Cependant, et comme souligné par Galwey (1995), alors que la stratégie de prélèvement constant paraît être satisfaisante pour assurer la représentation de la totalité de la gamme de variation des caractères utilisés pour la construction des classes, elle ne permet pas de représenter convenablement les formes les plus courantes, pouvant être intéressantes pour les programmes de sélection. La stratégie L, permettant une représentation plus équilibrée que les stratégies C et P, paraît être une bonne méthode d'échantillonnage aléatoire pour la constitution de core collection de sorgho.



Entre prélèvement aléatoire et sélectif, les différences ne sont pas significatives pour la représentation de la diversité globale. Cependant des caractéristiques spécifiques se rencontrent en fonction du type d'échantillonnage. La stratégie de prélèvement aléatoire après stratification semble influencer sur la distribution des caractères associés aux facteurs utilisés pour la stratification de la collection de base. La sélection basée sur la maximisation de la diversité d'un jeu de caractères (ici des caractères morpho-agronomiques) paraît ne pas avoir d'influence sur la variabilité des caractères phénotypiques non pris en compte dans la mise en oeuvre de la stratégie. La stratégie de sélection basée sur une large gamme d'informations morpho-agronomiques, agro-écologiques, culturelles... est la stratégie qui modifie le plus largement la distribution des caractères dans la core collection. Des hypothèses quand à l'origine de telles différenciations conduisent à envisager l'utilisation d'autres critères pour comparer ces trois stratégies de prélèvement.

Dans la partie suivante, nous analyserons l'impact des trois stratégies d'échantillonnage sur la diversité non prise en compte lors de l'échantillonnage et en particulier sur la diversité moléculaire.

**3<sup>ÈME</sup> PARTIE - IMPACT DE DIFFÉRENTS TYPES DE  
PRÉLÈVEMENT  
SUR LA DIVERSITÉ NON ÉCHANTILLONNÉE**

### 3<sup>EME</sup> PARTIE - IMPACT DE DIFFÉRENTS TYPES DE PRÉLÈVEMENT SUR LA DIVERSITÉ NON ÉCHANTILLONNÉE

A partir des stratégies d'échantillonnage aléatoire et non-aléatoire précédemment développées, nous avons montré que les trois sous-ensembles obtenus L, PCS et T, se caractérisent par des indices de diversité phénotypique globale comparables entre eux et à celui de la collection de base. Cependant, les distributions des caractères morpho-agronomiques ainsi que les répartitions raciale et géographique ne sont pas équivalentes dans tous les cas entre les sous-ensembles et la collection de base. Il semble que le mode d'échantillonnage ait une incidence sur la nature des caractères modifiés. Afin de préciser la nature des effets d'échantillonnage (biais ou spécificité de la stratégie), nous nous sommes intéressés aux effets de telles stratégies sur la diversité moléculaire. Si les différences de distributions phénotypiques sont corroborées au niveau moléculaire, alors elles pourraient être le résultat d'un biais d'échantillonnage. A l'inverse, si les différences ne sont pas confirmées, l'hypothèse d'un effet dû au mode de prélèvement serait la plus probable.

Dans cette troisième partie, nous nous proposons d'étudier les impacts sur la diversité moléculaire d'échantillonnages basés sur la diversité morpho-agronomique. Dans un premier temps, nous comparons les trois stratégies d'échantillonnage vis-à-vis de leur capacité à retenir la diversité moléculaire. Nos résultats présentés dans le chapitre 1 ont fait l'objet d'une publication acceptée dans *Theoretical and Applied Genetics*. La comparaison entre diversité phénotypique et moléculaire nous a conduit, dans un second temps, à rechercher d'éventuelles associations entre marqueurs moléculaires, caractères morpho-agronomiques, groupes raciaux et origines géographiques. Nos résultats sont présentés dans le chapitre 2. Enfin, à partir d'un ensemble constitué par les accessions des trois sous-ensembles nous avons estimé les effets, sur la diversité phénotypique, d'une stratégie d'échantillonnage basée sur la diversité moléculaire (chapitre 3).



## **CHAPITRE 1 : Impact des stratégies d'échantillonnage aléatoire vs. non aléatoire sur la diversité moléculaire**

Nous avons choisi d'estimer le polymorphisme moléculaire de l'ensemble des accessions à l'aide de marqueurs microsatellites. L'intérêt de ce type de marqueur est lié au polymorphisme élevé généralement révélé et à la facilité d'emploi notamment lors d'une utilisation en multiplex (PCR à partir de plusieurs couples d'amorces pour l'amplification de différents locus simultanément).

Dans cette étude, nous avons pu utiliser 15 couples d'amorces définies par Brown et al. (1996). Quatorze des 15 locus ainsi mis en évidence sont répartis sur 9 des 10 groupes de liaisons du sorgho (Dean et al., 1999). Ce travail a été réalisé au laboratoire de l'USDA-ARS-PGRU à Griffin, GA (USA).

Malgré une allofécondation naturelle relativement restreinte - de 5 à 10% (Doggett, 1988) - chez certaines races, l'allogamie peut aller jusqu'à 30% (Ollitrault, 1987 ; Chantereau et Kondombo, 1994). Cette observation rend compte du polymorphisme intrapopulation relativement important révélé dans les variétés traditionnelles de sorgho (Ollitrault, 1987, pour les sorghos burkinabés ; Djè et al., 1998 et 1999, pour les sorghos marocains). S'agissant de la collection maintenue à l'ICRISAT, à chaque cycle de multiplication, environ 30 panicules autofécondées sont récoltées pour reconstituer le stock de semences. Ces cycles conduisent à la réduction du polymorphisme intra variétal, à la fixation d'allèles et à des modifications de la structure génétique. Dans le but de tenir compte de l'existence d'une variabilité intra accession, nous avons choisi de représenter chaque accession par l'ADN extrait d'un mélange d'une trentaine de plantules.

Nos analyses ont montré que le taux d'homogénéité des accessions est variable selon le locus considéré (55% à 98%). Cependant, ces taux sont relativement semblables entre les trois sous-ensembles. L'origine de cette hétérogénéité n'est pas toujours facilement identifiable. Plusieurs phénomènes peuvent en rendre compte soit indépendamment soit en combinaison : hétérozygotie résiduelle, population naturelle constituée d'un ensemble de génotypes ou encore mélange de graines lors de la constitution des stocks de semences. Compte tenu de ces observations, nous avons choisi, dans notre travail, de prendre en compte uniquement les accessions homogènes.

**Tableau III-1.1** : diversité moléculaire estimée avec les richesses alléliques (nombre total d'allèles, No) et l'indice de diversité génétique de Nei (DI =  $1 - \sum(p_i^2)$ , où  $p_i$  est la fréquence du  $i^{\text{ème}}$  allèle (Nei, 1973)) pour l'ensemble global (EG) et chaque sous-ensemble (PCS, L et T).

Locus	EG	PCS	L	T
	No. (DI)	No. (DI)	No. (DI)	No. (DI)
Sb4-15	9 (0,80)	8 (0,81)	9 (0,76)	9 (0,77)
Sb4-121	9 (0,84)	9 (0,84)	9 (0,82)	8 (0,84)
Sb4-32	33 (0,91)	21 (0,91)	26 (0,89)	27 (0,91)
Sb5-236	24 (0,86)	19 (0,87)	19 (0,81)	19 (0,88)
Sb6-342	14 (0,79)	14 (0,80)	14 (0,77)	13 (0,79)
Sb6-36	20 (0,85)	17 (0,84)	19 (0,83)	18 (0,86)
Sb1-1	28 (0,86)	22 (0,83)	20 (0,88)	21 (0,85)
Sb1-10	31 (0,82)	26 (0,89)	20 (0,69)	23 (0,82)
Sb5-256	7 (0,27)	5 (0,30)	7 (0,23)	5 (0,28)
Sb6-84	23 (0,86)	21 (0,84)	20 (0,88)	16 (0,82)
Sb6-42	23 (0,93)	22 (0,93)	21 (0,92)	19 (0,90)
Sb6-57	16 (0,81)	12 (0,81)	15 (0,81)	11 (0,80)
Sb6-34	11 (0,74)	11 (0,76)	9 (0,71)	8 (0,74)
Sb5-206	26 (0,93)	25 (0,93)	24 (0,92)	23 (0,93)
Sb4-72	15 (0,71)	9 (0,72)	12 (0,66)	11 (0,75)
total	289	241	244	231
moyenne	19,3 (0,80)	16,1 (0,81)	16,3 (0,77)	15,4 (0,80)

**Tableau III-1.2** : nombre d'allèles fréquents (A), d'allèles rares dans l'ensemble global, EG (B), rares dans l'ensemble global mais présent dans les sous-ensemble PCS, L et T (C). S'agissant des allèles fréquents, une lettre (a, b ou c) partagée par deux ensembles indique une différence non significative à  $p<0,05$ .

Locus	EG		PCS		L		T	
	(A) $\chi^2$	(B)	(A) $\chi^2$	(C)	(A) $\chi^2$	(C)	(A) $\chi^2$	(C)
Sb4-15	6 a	3	6 a	2	5 b	3	6 a	3
Sb4-121	7 a	2	7 a	2	7 a	2	8 a	1
Sb4-32	8 a	25	8 a	11	8 b	18	8 a	19
Sb5-236	5 a, b	19	5 a	13	5 b	14	6 a, b	14
Sb6-342	6 a	8	7 a	8	5 a	8	6 a	7
Sb6-36	5 a	15	5 a	12	4 a	14	4 a	12
Sb1-1	5 a, b	23	4 a	17	6 a, b	15	4 b	16
Sb1-10	3 a	28	3 b	23	3 c	17	3 a, b, c	20
Sb5-256	2 a	5	2 a	3	2 a	5	2 a	3
Sb6-84	6 a, b	17	6 a	15	7 b	12	6 a	10
Sb6-42	9 a, b	14	9 a	13	9 a, b	12	7 b	10
Sb6-57	4 a, b	12	6 a, b	8	4 b	11	4 a	7
Sb6-34	4 a	7	4 a	7	4 a	5	4 a	4
Sb5-206	7 a	19	7 a	18	8 a	17	7 a	15
Sb4-72	6 a, b	9	6 a	3	4 a, b	5	5 b	5
total	83	206	85	157	81	158	80	146
moyenne	5.5		5.7		5.4		5.3	

Nos résultats montrent un polymorphisme important (Tableau III-1.1); 289 allèles au total pour l'ensemble global (constitué de la réunion des trois sous-ensembles) avec de 7 à 33 allèles par locus et une diversité génétique forte (de 0,71 à 0,93 sauf pour le locus Sb5-256). Ces résultats indiquent d'une part des distributions assez équilibrées entre allèles fréquents à chaque locus et d'autre part, par extrapolation, une forte diversité génétique au sein de la collection de base.

Des résultats similaires (nombre d'allèles, diversité génétique) ont été obtenus pour les trois sous-ensembles. Ainsi, quelle que soit la stratégie utilisée, prélèvement aléatoire ou sélectif, basée sur des caractères morpho-agronomiques, la diversité moléculaire globale retenue est forte et ne semble pas affectée par le mode de prélèvement.

La comparaison des distributions alléliques à chaque locus entre les trois sous-ensembles, montre des différences significatives pour quelques locus (Tableau III-1.2). De façon intéressante, le sous-ensemble L diffère des deux autres sous-ensembles pour un jeu de trois locus. Ainsi une spécificité paraît caractériser ce sous-ensemble L. Dans la mesure où ce dernier diffère des deux autres sous-ensembles par les caractères associés aux critères pris en compte lors de la stratification de la collection de base, un tel résultat conduit à supposer l'existence d'associations entre les marqueurs moléculaires, et en particulier les trois locus, et les facteurs associés au photopériodisme.

D'une manière générale, la représentation des allèles rares (fréquence inférieure à 5%) est relativement élevée dans chaque sous-ensemble (de 146 à 158 allèles, Tableau III-1.2). Cependant, le sous-ensemble T se caractérise par une perte plus importante d'allèles rares dont 31 spécifiquement perdus dans le sous-ensemble T contre 15 et 19 dans les sous-ensembles L et PCS. Cette stratégie sélective basée sur une connaissance des sorghos acquises au cours de nombreuses années et prenant en compte la diversité apparente des caractères morpho-agronomiques et/ou raciaux, des utilisations, des origines géographiques... entraîne la perte d'un plus grand nombre d'allèles rares. Cette observation est en accord avec celles plus générales faites sur de nombreuses plantes et qui tendent à indiquer une certaine indépendance entre diversité génétique et diversité morpho-agronomique. Le sous-ensemble L semble devoir son taux élevé de variétés fixées (82% d'accessions homogènes en moyenne sur les 15 locus, contre 76% dans les deux autres sous-ensembles) à la sur-représentation de la race kafir caractérisée par une faible diversité morpho-



agronomique, enzymatique (Ollitrault et al., 1989) et moléculaire (Morden et al., 1989 ; Aldrich et al., 1992 ; Deu et al., 1994).

En conclusion, les trois stratégies d'échantillonnage, comparables pour leur diversité phénotypique globale le sont aussi pour la diversité moléculaire globale. Cependant, les différences entre les trois stratégies, perceptibles au niveau de la distribution de certains caractères morpho-agronomiques, sont confirmées avec les marqueurs moléculaires. Parmi les trois sous-ensembles, la différenciation du sous-ensemble L par les trois mêmes locus, suggère un phénomène non-aléatoire. L'hypothèse d'une distinction des distributions phénotypiques et alléliques due au simple biais de prélèvement paraît peu probable. La stratégie d'échantillonnage mise en œuvre jouerait un rôle important. Les critères pris en compte pour la stratification de la collection de base précédant le prélèvement aléatoire seraient à l'origine de la différenciation du sous-ensemble L.

La distinction du sous-ensemble L vis-à-vis des deux autres sous-ensembles par des caractères associés au photopériodisme, mais également par trois locus communs nous amène à étudier d'éventuelles associations. Ainsi, nous nous proposons dans le chapitre suivant d'étudier les associations entre les marqueurs microsatellites et les caractères morpho-agronomiques, et les groupes raciaux.

L'échantillonnage établi sur les caractères morpho-agronomiques semblent ne pas affecter la diversité moléculaire globale. Quel est l'impact d'un prélèvement constitué à partir des marqueurs moléculaires sur la diversité phénotypique ? L'échantillonnage aléatoire stratifié modifie la distribution allélique à certains locus. Quel serait l'effet d'un prélèvement prenant en compte ces locus sur les caractères structurant de la diversité dans la stratégie de prélèvement aléatoire ? Ces questions seront traitées dans le dernier chapitre de cette partie.

## CHAPITRE 2 : Recherche d'associations entre marqueurs

L'effet de la stratégie L sur la distribution allélique à certains locus (Sb4-15, Sb4-32 et Sb6-84), permet de distinguer de façon spécifique cette stratégie des deux autres. Une telle distinction laisse présager l'existence d'associations entre ces locus et les facteurs pris en compte dans cette stratégie d'échantillonnage. Des études d'associations entre ces marqueurs moléculaires, les caractères morpho-agronomiques et les groupes raciaux sont proposées dans ce chapitre.

### Matériel et Méthodes

#### *matériel*

Les 15 marqueurs microsatellites ont été étudiés pour les trois sous-ensembles (L, PCS et T). Seuls les allèles fréquents (>5% dans l'ensemble des accessions homogènes au locus considéré) et les accessions homogènes ont été retenus. Ainsi, les accessions ont pu être codées comme des génotypes homozygotes.

Les données utilisées pour la recherche d'associations entre marqueurs et caractères morpho-agronomiques sont celles obtenues dans nos conditions expérimentales et décrites dans la partie I.

#### *méthodes*

A partir des fréquences marginales d'un tableau de contingence 2x2, le test de Fisher calcule la probabilité *exacte*, sous l'hypothèse nulle que les deux facteurs ne sont pas corrélés dans la population, d'obtenir la distribution des fréquences observée. Ces tests d'indépendance ont été effectués avec le logiciel Genepop version 3.1b (Raymond et Rousset, 1997). La procédure des chaînes de Markov, basée sur la réitération de comparaisons de nombreux tableaux deux à deux, est utilisée pour obtenir une valeur non biaisée de p. Le seuil de signification pour lequel nous choisissons de refuser l'hypothèse d'indépendance est fixé à  $\alpha = 5\%$ . Cependant, comme un grand nombre de comparaisons sont testées, le seuil est ainsi modifié par la correction réalisée par la méthode séquentielle de Bonferroni pour un test multilocus (Rice, 1989, cité dans Carlier et al., 1996). Ainsi, pour tester l'indépendance de N locus,  $\alpha' = 0,05/[N(N-1)/2]$  et dans notre étude  $\alpha' = 4,7.10^{-4}$ .

Nous avons ainsi testé les associations de chacun des trois locus considérés avec les autres locus microsatellites, les caractères morpho-agronomiques et les origines géographiques.

## Résultats et Discussion

Au total, beaucoup plus d'associations entre marqueurs moléculaires ont été détectées dans le sous-ensemble L (25) que dans PCS (6) et dans T (4). Quatorze de ces associations impliquent au moins un des trois locus dont la distribution a été modifiée dans l'échantillonnage L (Sb4-15, Sb4-32, Sb6-84) (Tableau III-2.1). De manière globale, les associations entre locus sont plus nombreuses dans le sous-ensemble L (12 contre 4 et 1, respectivement dans les sous-ensembles PCS et T). Les trois locus mentionnés ci-dessus, sont tous trois fortement associés dans le sous-ensemble L. En revanche ces associations ne sont pas toutes maintenues dans les deux autres sous-ensembles. Le locus Sb6-57 est aussi associé au jeu des trois locus dans le sous-ensemble L et une de ces associations (Sb4-15/Sb6-57) est aussi maintenue dans les deux autres sous-ensembles. Un autre locus, Sb4-72 présente aussi des associations avec les trois locus considérés. Une seule association (Sb4-15/Sb4-72) a été trouvée à la fois dans les sous-ensemble L et PCS. Il semble donc que les associations entre marqueurs soient dépendantes pour une bonne majorité des stratégies d'échantillonnage mise en oeuvre. Ces associations résultant d'effet fondateur peuvent être maintenues par la structuration raciale des sorghos cultivés, ou par des associations génétiques - à considérer avec la position des marqueurs sur les groupes de liaison ou par l'existence d'associations résultant de liaisons indirectes (marqueurs proches de gènes en déséquilibre de liaison seraient transmis ensembles par effet "auto-stop"). Les informations disponibles sur la cartographie de ces locus microsatellites indiquent que Sb4-15 et Sb4-32 sont bien sur le même groupe de liaison, mais à quelle distance ? Par contre, les locus Sb6-84, Sb6-57 et Sb4-72 sont sur des groupes de liaison différents.

Pour les accessions du sous-ensemble L, des associations entre marqueurs et caractères morpho-agronomiques sont aussi mises en évidence. Les locus Sb4-15 et Sb6-84 sont associés aux caractères du cycle végétatif à la saison kharif et à la race, le locus Sb4-32 montre de fortes associations pour la hauteur de plante à la saison kharif, la texture de l'albumen et la race. Pour le caractère du cycle végétatif à la saison kharif, les variétés précoces et/ou petites (FLK1 et PHTK1) tendent à montrer une



**Tableau III-2.2** : distribution allélique des locus Sb4-15, Sb4-32 et Sb6-84 en fonction de trois caractères morpho-agronomiques et de l'appartenance raciale des accessions analysées du sous-ensemble L. Seuls les allèles fréquents dénommés par leur taille en pb ont été pris en compte.

‡ regroupe les intermédiaires DB, GB, KB, CB, KC, DG, DK, GK

Sb4-15								Sb6-84							Sb4-32									
	118	122	124	126	132	134	Total		182	184	186	194	198	Total		173	177	179	181	189	191	193	197	Total
PHTK1	0	2	7	26	11	1	47	PHTK1	4	4	8	20	9	45	PHTK1	18	4	3	5	4	1	6	3	44
PHTK2	0	6	6	15	7	1	35	PHTK2	3	6	3	10	8	30	PHTK2	8	1	2	3	3	2	8	3	30
PHTK3	1	4	1	6	13	1	26	PHTK3	8	6	2	0	0	16	PHTK3	2	1	2	1	3	1	8	2	20
PHTK4	6	7	5	1	18	8	45	PHTK4	14	12	0	4	3	33	PHTK4	8	8	4	1	10	3	0	0	34
Total	7	19	19	48	49	11	153	Total	29	28	13	34	20	124	Total	36	14	11	10	20	7	22	8	128

Sb4-15								Sb6-84							Sb4-32									
	118	122	124	126	132	134	Total		182	184	186	194	198	Total		173	177	179	181	189	191	193	197	Total
FLK1	0	3	11	28	7	0	49	FLK1	4	6	9	21	12	52	COR1	2	3	1	0	0	0	0	0	6
FLK2	0	7	4	11	11	4	37	FLK2	5	8	1	8	4	26	COR2	6	4	2	1	3	1	2	0	19
FLK3	0	2	1	2	24	3	32	FLK3	10	7	2	1	1	21	COR3	24	1	2	7	5	1	12	4	56
FLK4	7	7	3	7	7	4	35	FLK4	10	7	1	4	3	25	COR4	5	5	4	2	10	5	2	1	34
Total	7	19	19	48	49	11	153	Total	29	28	13	34	20	124	COR5	0	1	2	0	2	1	6	3	15
															Total	37	14	11	10	20	8	22	8	130

Sb4-15								Sb6-84							Sb4-32									
	118	122	124	126	132	134	Total		182	184	186	194	198	Total		173	177	179	181	189	191	193	197	Total
bicolor	0	2	3	0	8	0	13	bicolor	4	3	0	2	1	10	bicolor	1	0	1	0	1	1	0	0	4
caudatum	0	5	0	6	7	4	22	caudatum	2	11	5	2	1	21	caudatum	2	1	1	2	5	4	8	3	26
durra	0	1	1	2	11	1	16	durra	5	1	0	1	0	7	durra	3	0	2	3	1	1	1	0	11
guinea	6	1	3	2	4	2	18	guinea	7	4	1	2	1	15	guinea	6	6	1	0	5	0	0	0	18
kafir	0	0	4	16	0	0	20	kafir	0	0	0	11	6	17	kafir	12	0	0	0	0	0	3	1	16
DC	0	5	1	2	5	0	13	DC	1	1	2	2	2	8	DC	1	0	3	1	3	0	2	1	11
GC	0	2	2	10	9	4	27	GC	5	7	2	8	4	26	GC	6	6	0	2	3	1	4	1	23
autres ‡	1	4	5	11	5	0	26	autres ‡	5	3	3	6	5	22	autres ‡	6	1	3	2	2	1	5	2	22
Total	7	20	19	49	49	11	155	Total	29	30	13	34	20	126	Total	37	14	11	10	20	8	23	8	131

distribution allélique avec un pic de distribution au locus Sb4-15 pour l'allèle 126 (57% pour les précoces et 55% pour les petites) et au locus Sb6-84 pour l'allèle 194 (40% pour les précoces et 44% pour les petites) (Tableaux III-2.2). Au locus Sb4-32 et pour les variétés de petite taille, la distribution allélique est en faveur de l'allèle 173. Pour ces trois locus, les allèles majeurs identifiés chez les variétés précoces et petites, essentiellement associés aux variétés photoinsensibles, sont identiques à ceux trouvés chez les sorghos kafir. Pour ces locus, la distribution allélique montre que 80% des variétés kafir sont caractérisés par l'allèle 126 au locus Sb4-15, 65% des kafir par l'allèle 194 au locus Sb6-84, et 75% des kafir par l'allèle 173 au locus Sb4-32. Pour les variétés plus tardives et/ou de plus grande taille, des allèles fréquents sont mis en évidence, allèle 132 pour le locus Sb4-15, allèle 182 et 184 pour le locus Sb6-84 et allèle 189 pour le locus Sb4-32. Ces allèles aux locus Sb4-15 et Sb6-84 sont ceux identifiés majoritairement chez les sorghos durra et caudatum (dans la limite des effectifs). Des associations entre les phénotypes extrêmes du caractère de vitrosité des grains et le locus Sb4-32 sont aussi mises en évidence. Ainsi, les sorghos à grains vitreux (COR1) sont ceux qui ont les allèles possédant les plus petites tailles (de 173 à 179 pb), alors que les variétés à grains farineux (COR5) possèdent en majorité les allèles 193 à 197. Ce caractère est fortement associé aux variétés guinea, ces dernières étant essentiellement caractérisées par des grains vitreux (85%) et majoritairement caractérisées avec l'allèle 173.

Les locus affectés par la stratégie L se sont révélés être fortement associés aux caractères morpho-agronomiques, étant eux mêmes associés au photopériodisme. Ces trois locus ont permis de révéler une structuration raciale et notamment une distinction des variétés de la race kafir qui sont largement représentées par cette stratégie. Ainsi, les effets de l'échantillonnage logarithmique se répercutant jusqu'au niveau de la distribution des allèles, le choix des critères à prendre en compte pour la structuration de la collection de base n'est pas sans conséquences. Le biais introduit par le rééquilibrage des groupes dans l'échantillon L conduit à des distorsions au niveau de l'ensemble des caractères associés directement ou non à ceux pris en compte pour la constitution des groupes initiaux dans la collection de base. La question que nous pouvons alors nous poser est celle de l'impact sur la diversité morpho-agronomique d'un échantillonnage constitué à partir des locus associés ou non aux facteurs utilisés pour la hiérarchisation de la collection de base.

### **CHAPITRE 3 : Impact sur la diversité morpho-agronomique d'un échantillonnage constitué à partir de la diversité moléculaire**

Dans les chapitres précédents, nous avons pu noter que la sélection basée sur des caractères morpho-agronomiques n'influence pas la diversité moléculaire globale. De plus, nous avons pu vérifier qu'une stratégie de prélèvement aléatoire après stratification induisant des modifications dans la distribution des caractères morpho-agronomiques, peut aussi affecter les locus associés aux critères pris en compte lors de la hiérarchisation de la collection de base.

Peut-on utiliser les marqueurs moléculaires pour la constitution de core collections ? Quel est l'impact sur la diversité morpho-agronomique d'un échantillonnage constitué à partir de la diversité moléculaire ? Cet impact est-il lié ou non aux caractéristiques spécifiques des marqueurs utilisés ? Autrement dit, aura-t-on des résultats similaires quels que soient les locus considérés ?

Pour apporter des éléments de réponse à ces questions, nous nous proposons d'établir une sélection suivant une stratégie d'échantillonnage basée sur des données moléculaires. Nous avons vu que la stratégie PCS permet une bonne représentation de la diversité (variances inchangées, redondances fortement réduites) et favorise la sélection des individus qui contribuent au plus à la diversité. Cette stratégie utilisant à l'origine les caractères quantitatifs peut également être utilisée pour des variables qualitatives de type données moléculaires (Barcelos, 1998 ; Hamon et al., 1998). Cette méthode a donc été utilisée sur nos données de caractérisation moléculaire obtenues à partir des marqueurs microsatellites.

#### **Matériel et Méthodes**

##### *matériel*

A partir de la réunion des trois sous-ensembles L, PCS et T, nous avons constitué deux ensembles initiaux comportant toutes les accessions sans données manquantes pour deux jeux de cinq locus.

##### *méthodes*

Deux sous-échantillonnages sont réalisés. Un premier (Cqall) effectué sur un ensemble initial de 453 accessions porte sur cinq locus dont les trois associés aux



**Tableau III-3.1** : comparaison des distributions des caractères morpho-agronomiques entre l'ensemble initial et le sous-échantillon Cqal1. **A.** caractères quantitatifs - moyenne (M±se) et variance - et tests d'homogénéité de variances (Levene) et de comparaison de moyennes (Newman-Keuls) - **B.** caractères qualitatifs (test de chi-2).

A

	Cqal1			
	ensemble initial		sous-échantillon	
	M±se	Variance	M±se	Variance
FLR	63,54±0,48	105,53	60,91±1,42	90,58
FLK	96,13±1,55	1062,99	100,52±5,01	1103,42
PHTR	244,26±2,39	2578,90	244,89±5,75	1488,06*
PHTK	317,76±4,53	9052,68	323,07±11,4	5714,20*
BT	1,1±0,02	0,13	1,09±0,07	0,22
PEDEX	11,87±0,41	74,94	12,67±1,35	81,64
EHLG	25±0,41	74,80	23,89±1,5	101,74
EHWD	10,34±0,21	19,35	10,76±0,71	22,64
GRS	2,99±0,03	0,54	2,87±0,12	0,64
SWT	2,9±0,05	1,27	2,67±0,17	1,34

B

Cqal1	
	$\chi^2$
PIG	ns
MRC	ns
EHCS	ns
GLC	ns
COV	**
GRC	**
COR	ns
SC	**

\*, \*\* : différence significative à 5% et 1%

**Tableau III-3.2** : comparaison des distributions raciale et géographique et entre ensemble initial et sous-échantillon Cqal1

	Cqal1			
	ensemble initial		sous-échantillon	
	nombre	%	nombre	%
bicolor	31	6,84	7	15,56
caudatum	59	13,02	9	20,00
durra	78	17,22	6	13,33
guinea	66	14,57	4	8,89
kafir	34	7,51	2	4,44
CB	23	5,08	4	8,89
DB	18	3,97	4	8,89
DC	35	7,73	4	8,89
GC	61	13,47	5	11,11
GB - GD	12	2,65	0	
intermédiaires kafir	36	7,95	0	
<b>Total races</b>	<b>453</b>		<b>45</b>	
Amérique du nord	24	5,30	0	
Amérique latine	2	0,44	0	
Arabie	1	0,22	1	2,22
Afrique du nord et Europe du sud	1	0,22	0	
Afrique orientale	127	28,04	17	37,78
Afrique centrale	60	13,25	3	6,67
Afrique occidentale	112	24,72	12	26,67
Afrique australe	66	14,57	1	2,22
sous-continent indien	49	10,82	8	17,78
Extrême orient et Océanie	11	2,43	3	6,67
<b>Total régions</b>	<b>453</b>		<b>45</b>	

critères du photopériodisme (Sb4-15, Sb4-32, Sb6-36, Sb6-84 et Sb6-57). Le deuxième sous-échantillonnage (Cqal2) effectué sur 410 accessions est obtenu à partir de cinq autres locus pour lesquels les distributions dans les sous-ensembles L, PCS et T sont similaires (Sb4-121, Sb6-342, Sb5-256, Sb6-34 et Sb5-206).

#### *échantillonnage*

La sélection des accessions se fait suivant une analyse factorielle des correspondances, réalisée sur les tableaux disjonctifs complets ("présence/absence"). Les distances calculées entre les accessions sont basées sur les estimations des chi-2, et sont pondérées par les racines carrées des valeurs propres des facteurs correspondants. Les accessions sont ensuite sélectionnées successivement en fonction de leur contribution à la somme des carrés généralisée, afin de maximiser la diversité retenue.

#### *analyse des données*

Les deux sous-échantillons Cqal1 et Cqal2, représentant chacun 10% de la taille de leur ensemble initial, sont étudiés vis-à-vis de leur variabilité morpho-agronomique. A partir des données collectées dans nos conditions expérimentales (cf. partie I), les distributions des caractères morpho-agronomiques de chaque sous-échantillon et de son ensemble initial sont comparées par des tests d'homogénéité de variance (test de Levene), de comparaison de moyennes (test de Newmann-Keuls) et des tests de chi-2.

### **Résultats**

Pour l'ensemble des caractères morpho-agronomiques, appartenance raciale et distribution géographique, nous avons tout d'abord montré que les deux ensembles initiaux sont semblables.

En prélevant 10% de chaque ensemble initial, nous avons obtenu des sous-échantillons retenant 45% et 41% de la variabilité moléculaire totale (respectivement pour Cqal1 et Cqal2). Dans le sous-échantillonnage Cqal1 (les locus pris en compte sont associés au photopériodisme), certains caractères morpho-agronomiques ont une distribution modifiée (Tableau III-3.1). Les variances des caractères de hauteur de plante à la saison rabi (PHTR) et à la saison kharif (PHTK), sont significativement réduites dans le sous-échantillon. Les caractères de recouvrement des glumes (COV), couleur des grains (GRC) et présence/absence d'une sous-couche brune (SC) ont une distribution significativement modifiée par l'échantillonnage. Les variétés retenues dans ce sous-échantillon sont principalement caractérisées par des grains très

**Tableau III-3.3** : comparaison des distributions des caractères morpho-agronomiques entre l'ensemble intial et le sous-échantillon Cqal2. **A.** caractères quantitatifs - moyenne (M±se) et variance - et tests d'homogénéité de variance (Levene) et de comparaison de moyenne (Newman-Keuls) - **B.** caractères qualitatifs (test de chi-2).

A					B	
	Cqal2					Cqal2
	ensemble initial		sous- échantillon			$\chi^2$
	M±se	Variance	M±se	Variance		
FLR	62,94±0,47	92,70	61,95±1,62	107,70	PIG	ns
FLK	93,05±1,64	1074,50	97,54±5,66	1247,36	MRC	ns
PHTR	242,85±2,49	2550,29	242,56±7,6	2368,90	EHCS	ns
PHTK	312,69±4,89	9595,76	320,77±13,89	7520,45	GLC	ns
BT	1,1±0,02	0,13	1,27±0,11	0,50**	COV	ns
PEDEX	11,84±0,42	71,22	12,15±1,71	119,38**	GRC	*
EHLG	24,58±0,42	71,39	24,71±1,18	57,41	COR	ns
EHWD	10,24±0,21	18,77	10,8±0,85	29,61	SC	ns
GRS	2,99±0,04	0,52	2,85±0,13	0,67		
SWT	2,88±0,06	1,26	2,65±0,17	1,16		

\*, \*\* : différence significative à 5% et 1%

\*, \*\* : différence  
significative à 5% et 1%

**Tableau III-3.4** : comparaison des distributions raciales et géographiques entre ensemble initial et sous-échantillon

	Cqal2			
	ensemble initial		sous-échantillon	
	nombre	%	nombre	%
bicolor	19	4,59	7	17,07
caudatum	65	15,70	7	17,07
durra	61	14,73	7	17,07
guinea	58	14,01	3	7,32
kafir	33	7,97	0	
CB	24	5,80	3	7,32
DB	20	4,83	4	9,76
DC	28	6,76	5	12,20
GC	55	13,29	5	12,20
GB - GD	11	2,66	0	
intermédiaires kafir	40	9,66	0	
<b>Total races</b>	<b>414</b>		<b>41</b>	
Amérique du nord	13	3,15	2	5,00
Amérique latine	2	0,48	1	2,50
Arabie	1	0,24	1	2,50
Afrique orientale	120	29,06	14	35,00
Afrique centrale	58	14,04	5	12,50
Afrique occidentale	91	22,03	7	17,50
Afrique australe	64	15,50	1	2,50
sous-continent indien	38	9,20	4	10,00
Extrême orient et Océanie	26	6,30	6	15,00
<b>Total régions</b>	<b>413</b>		<b>40</b>	



recouverts par les glumes. En majorité ce sont des sorghos à grains bruns ou gris et/ou présentant une sous-couche brune. La race bicolor est plus de deux fois plus fréquente dans le sous-échantillon que dans l'ensemble initial (Tableau III-3.2). L'essentiel du matériel de ce sous-échantillon est originaire d'Afrique orientale (38%) et d'Afrique occidentale (27%), et la représentation des variétés sud-africaines est significativement réduite, passant d'environ 15% dans le set initial à 2% dans le sous-échantillon.

L'échantillonnage établi sur un jeu de locus non affectés par les stratégies de prélèvement basé sur les caractères morpho-agronomiques (Cqal2) se différencie de son ensemble initial par une augmentation significative de la variance du nombre de talles racinaires (BT) et de la longueur du pédoncule (PEDEX) (Tableau III-3.3). La distribution pour le caractère de coloration des grains est aussi significativement différente, avec une représentation en faveur des variétés à grains bruns. En ce qui concerne la distribution des races, la fréquence des variétés bicolor est là encore bien plus élevée dans le sous-échantillon Cqal2, que dans l'ensemble initial, respectivement 17,1% et 4,6% (Tableau III-3.4). Dans ce sous-échantillon, les sorghos de la région Extrême-orient et Océanie sont plus de deux fois plus représentés que dans l'ensemble initial (15% contre 6,3%). A l'inverse, la représentation des sorghos d'Afrique du Sud est très fortement réduite dans le sous-ensemble.

## Discussion

D'une façon générale, les stratégies établies sur les marqueurs microsatellites semblent affecter les distributions de quelques caractères morpho-agronomiques, races et des origines géographiques. Le caractère morpho-agronomique modifié par les deux échantillonnages présente des caractéristiques particulières (coloration des grains) qui sont majoritaires chez certaines races. Ainsi, les variétés à grains bruns, dont la représentation est favorisée par les deux échantillonnages, appartiennent essentiellement aux races bicolor (76%) et caudatum (61%). Dans les deux sous-échantillons obtenus, les variétés bicolor sont plus de deux fois plus représentées que dans les ensembles initiaux respectifs. La race bicolor, considérée proche de la forme primitive des sorghos cultivés, révèle une grande variabilité morpho-agronomique mais aussi une importante diversité génétique notable avec les marqueurs enzymatiques, RFLPs nucléaires, et microsatellites. Ainsi, quel que soit le jeu de marqueurs moléculaires pris en compte pour l'échantillonnage, la représentation de

cette race hautement diversifiée devrait être favorisée. A l'inverse, la race kafir est assez peu représentée dans les deux sous-échantillons. Un tel biais peut être expliqué par la base génétique étroite de cette race dont la diversification est récente. Par conséquent, et compte tenu du fait que la race kafir représente une part importante des sorghos cultivés en Afrique australe (36%), la représentation de cette région (de faible diversité génétique) est diminuée (environ 15% dans les deux sets initiaux et 2% dans les deux sous-échantillons). A l'inverse, les sorghos cultivés en Afrique orientale, présentant une diversité génétique élevée, sont très bien représentés dans les deux échantillonnages.

Malgré ces tendances partagées entre les deux sous-échantillonnages, les impacts du prélèvement sur la distribution des caractères peuvent être différents selon les locus pris en compte. Dans tous les cas, la stratégie PCS basée sur la maximisation de la diversité va conduire à la sélection en premier lieu des individus présentant le maximum d'allèles rares et de combinaisons alléliques originales. La différence entre les deux sous-échantillons Cqal1 et Cqal2 s'opère sur la distribution raciale des allèles rares aux locus considérés. Dans le 1<sup>er</sup> jeu de locus utilisé, environ 1/10<sup>ème</sup> des allèles rares sont apportés par les kafir (8/83) dont quelques uns apportés de manière spécifique. De même, de nombreux allèles rares et combinaisons alléliques originales sont apportés par les bicolor et caudatum. Cette sélection ne favorise ni les guinea ni les durra dont la représentation passe respectivement de 14,6% à 9% et de 17,2 à 13,3%. Par contre-coup, les très grandes plantes (surtout des guinea) ne sont pas représentées, conduisant ainsi à une réduction de la variance de PHTR et une moyenne plus petite que l'ensemble initial.

Dans le 2<sup>ème</sup> jeu de locus, quatre allèles sur 44 rares sont apportés par des kafir mais sont présents aussi dans d'autres races. Des combinaisons alléliques originales sont favorables à une bonne représentation des bicolor des intermédiaires DB et DC. Les guinea ou les kafir ne possédant pas d'allèles rares spécifiques et/ou de combinaison allélique originale sont sous-représentés, voire absent de la sélection. La sélection vis-à-vis de ce jeu de locus induit un prélèvement aléatoire des allèles aux autres locus et finalement une bonne représentation pour un grand nombre de caractères.

Si on reprend l'ensemble initial de 453 individus (à l'origine de Cqal1) et que l'on applique une sélection PCS uniquement sur les allèles fréquents (>5%), on constate que dans un sous-ensemble de 10% les bicolor ont une représentation pratiquement inchangée par rapport à l'ensemble initial, alors que les guinea passent de 14,6% à

26,6%. Cette différence de représentation raciale a pour conséquence une augmentation de la variance de la hauteur de plante aux deux saisons et un déplacement de la moyenne vers les plus grandes plantes.

En conclusion, à priori, cinq locus microsatellites sont suffisants pour avoir une assez bonne représentation de la diversité morpho-agronomique, qu'ils soient structurants ou non. Par contre, il semble qu'il faille davantage considérer les types d'allèles (fréquents, rares et fréquents) à prendre en compte. L'utilisation des allèles rares (très nombreux avec les locus microsatellites ; 206/289 pour 15 locus étudiés cf. partie III) peut induire des modifications de la distribution de certains caractères en raison de la présence aléatoire de ces allèles rares dans telle ou telle catégorie d'accessions.

D'autre part, la stratégie utilisée (PCS) joue aussi un rôle important dans l'observation des modifications induites par l'échantillonnage. L'application d'une stratégie de prélèvement aléatoire devrait permettre de s'approcher de la situation observée uniquement en considérant PCS sur allèles fréquents. Les deux stratégies de prélèvement : aléatoire avec l'ensemble des allèles, et PCS sur allèles fréquents, se différencieront probablement par la diversité retenue en faveur à priori pour PCS d'une plus grande richesse en terme de combinaisons alléliques.

Ces résultats amènent à deux réflexions : faut-il accorder plus d'importance aux combinaisons alléliques ou aux allèles pris "individuellement" ? Les microsatellites ont d'emblée été considérés très intéressants car marqueurs offrant le plus fort degré de polymorphisme. Ne sont-ils pas, dans certains cas, trop polymorphes pour justement appréhender la situation ?

Dans cette partie, plusieurs points sur les relations entre diversité phénotypique et diversité moléculaire ont été abordés. Nous avons pu montrer qu'indépendamment de la stratégie d'échantillonnage utilisée (L, PCS et T), un prélèvement basé sur la diversité morpho-agronomique ne semble pas affecter la diversité moléculaire globale. En revanche, les locus Sb4-15, Sb4-32 et Sb6-84 présentent une distribution allélique spécifique dans le sous-ensemble L. Ce résultat nous a permis de suggérer l'existence dans ce sous-ensemble d'éventuelles associations entre ces trois locus et les caractères de photopériodisme utilisés pour hiérarchiser la collection de base. Des associations ont été mises en évidence dont celles faisant intervenir les trois locus et les hauteurs de plante, et dates de floraison.



Nous nous sommes ensuite proposés d'étudier les effets sur la diversité morpho-agronomique d'un échantillonnage à partir de la diversité moléculaire. Nous avons comparé des stratégies prenant en compte ou non les trois locus associés au photopériodisme (respectivement Cqal1 et Cqal2). Les résultats obtenus indiquent une bonne représentation de la diversité morpho-agronomique par un échantillonnage basé sur la diversité moléculaire. Cependant, le choix de la stratégie (sélective de type PCS ou aléatoire) et le type d'allèles à prendre en compte (tous les allèles ou uniquement les allèles fréquents) sont importants à considérer. Si les marqueurs moléculaires sont généralement considérés comme neutres, la nature de ces marqueurs (microsatellites, RFLPs) n'est peut être pas neutre pour l'application d'une stratégie d'échantillonnage basée sur la diversité moléculaire.

## **CONCLUSION - DISCUSSION GENERALE**

## CONCLUSION - DISCUSSION GENERALE

L'utilisation efficace des ressources génétiques implique nécessairement la connaissance de la diversité génétique contenue dans ces ressources génétiques. Sur ce point, les informations de la base de données de l'ICRISAT acquises régulièrement depuis ces 25 dernières années, nous ont permis (après validation des données) d'effectuer une étude descriptive et détaillée de l'ensemble des variétés traditionnelles. Dans la collection, les origines géographiques des sorghos cultivés sont assez bien représentées, avec toutefois des exceptions comme la Chine et certains pays d'Afrique. La représentation des races est en accord avec leur importance en culture. Les races caudatum, durra, guinea, dura-caudatum et guinea-caudatum sont très fréquentes dans la collection, alors qu'il y a peu de variétés de race bicolor, malgré leur forte diversité génétique. La diversité phénotypique globale de la collection est importante. Les caractéristiques morpho-agronomiques des races, l'importance de la diversité génétique des régions considérées sont en règle générale en accord avec le schéma de domestication des sorghos cultivés, la taxonomie, et les observations faites par d'autres auteurs (Chantereau et al., 1989 ; Ollitrault et al., 1989 ; Appa Rao et al., 1996 ; Teshome et al., 1997 ; Ayana et Bekele, 1999 ; Deu et al. 1999). Néanmoins, chez certaines races, nous avons noté la présence de quelques variétés alliant des caractéristiques morpho-agronomiques tout à fait originales et pouvant présenter un intérêt pour des programmes de sélection variétale. Ainsi, la majorité des variétés caudatum présente des grains farineux, mais quelques variétés à grains vitreux ont été identifiées. Cette caractéristique est très intéressante pour des programmes de sélection axés sur l'obtention de variétés améliorées alliant les caractères de productivité des caudatum et ceux de qualité des grains des guinea (Chantereau et al., 1998 ; Rami et al., 1998). Les caudatum sont traditionnellement considérés comme des sorghos moyennement photosensibles. Au sein de cette race, Deu et al. (1999) ont mis en évidence l'existence de deux groupes morphologiques  $C_1$  et  $C_2$ . Selon ces auteurs, la différenciation des groupes aurait pour origine la sensibilité à la photopériode. Notre étude confirme l'existence d'une large gamme de variation au photopériodisme chez les caudatum, avec notamment quelques variétés très photosensibles originaires du Cameroun, du Kenya ou de l'Ouganda. Cette gamme de variation phénotypique notée aussi au niveau de la forme de la panicule chez les caudatum mais aussi les guinea-caudatum et durra-caudatum, leur confère un intérêt



particulier pour les programmes de sélection variétale. En effet, la compacité de la panicule est fortement corrélée aux conditions hygrométriques. Plus la panicule est compacte, plus la variété est adaptée aux conditions sèches

Bien que de manière ponctuelle il soit très probable que la collection maintenue à l'ICRISAT Asian Center (IAC) soit moins riche qu'une collection constituée à l'échelle d'un pays ou d'une région, notre étude montre que cette collection est une source de diversité génétique importante. Pourtant, dans la pratique, cette collection a-t-elle était bien utilisée par le passé, le sera-t-elle davantage dans l'avenir ?

Jusqu'à ces dernières années, pour les utilisateurs potentiels hors ICRISAT, la collection de l'IAC se résumait à un catalogue contenant des informations passeport minimales. Pour des raisons d'ordre technique, les informations concernant les évaluations étaient alors difficiles d'accès (fichiers de données importants, difficilement manipulables et encore plus difficilement diffusables). Par conséquent, cette collection a pu être sous-exploitée mais aussi et sans doute mal exploitée (diversité mal connue). Il est difficile de demander ce dont l'existence même est ignorée.

Depuis ces cinq dernières années, l'accès aux données ICRISAT via le réseau électronique est possible. Dans ce contexte, notre étude devrait permettre de susciter de l'intérêt pour la collection maintenue à l'IAC. Dans les années à venir, la consultation et l'utilisation des données devraient elles aussi augmenter, favorisant la valorisation des ressources génétiques et justifiant grandement la préservation d'une telle collection.

Ce regain d'intérêt pour cette collection pourrait avoir des conséquences sur sa gestion et sur la mise en place de priorités pour les évaluations futures.

La constitution de core collections est un moyen de définir ces priorités et de concentrer les efforts sur un ensemble de taille réduite (en général 10% de la collection de base) ; différentes stratégies d'échantillonnage ont été élaborées au cours de cette décennie. Dans notre étude, nous avons vu que la collection peut être hiérarchisée, en fonction de la sensibilité à la photopériode ; les quatre groupes obtenus étant inégaux en taille. Dans ces conditions, un prélèvement proportionnel (P) dans chaque groupe, conduit à une représentation assez fidèle de la collection de base. Par contre les prélèvements constant (C) et logarithmique (L) induisent des modifications plus ou moins importantes de la distribution des caractères. La comparaison des stratégies de prélèvement aléatoire (L) vs. non aléatoire

[maximisation de la diversité (PCS) et sélection empirique des accessions (T)] a montré que la core collection PCS possède des distributions de caractères assez semblables à celles de la collection de base. A l'inverse la core collection T diffère en tous points de la collection de base. La core collection L quant à elle, diffère de la collection de base par des caractères plus ou moins fortement liés au photopériodisme. Ces différences de comportement des trois core collections sont davantage liées à la structure de l'ensemble initial à échantillonner (hiérarchisé ou non) qu'à la méthode d'échantillonnage (aléatoire vs. non aléatoire). Sur ce point, chez le palmier à huile, Barcelos (1998) conclut à l'intérêt d'effectuer un échantillonnage non-aléatoire type PCS après stratification de la collection initiale. Ce faisant, il favorise l'échantillonnage de la diversité intragroupe tout en préservant la diversité intergroupe. Lorsque l'échantillonnage est réalisé à partir de la collection initiale non stratifiée, la diversité intergroupe est favorisée au détriment de la diversité intragroupe.

Ces résultats indiquent qu'il est important de connaître et de prendre en compte la structuration de la diversité. Cependant, il est tout aussi important de bien considérer la pertinence des facteurs pris en compte pour structurer cette diversité. Dans notre étude, la structuration vis-à-vis de la photopériode nous semble plus pertinente qu'une structuration uniquement raciale, car prenant en compte des caractéristiques de comportement vis-à-vis des conditions environnementales.

Le choix de la stratégie à utiliser pour la constitution de core collections dépend des objectifs fixés, et de la disponibilité ou non de données d'évaluation. Maximiser la représentation de la diversité de la collection de base est un objectif intéressant à atteindre. La sélection empirique ne peut pas être recommandée dans la mesure où elle nécessite une connaissance du matériel acquise après de longues années d'expérience. La sélection PCS peut être intéressante, mais la nature des caractères (morphologiques ou moléculaires) pris en compte pour l'application de cette stratégie est importante à considérer. L'utilisation de données quantitatives permet de calculer des distances de type euclidiennes. Dans des analyses multivariées, celles-ci offrent l'avantage de ne pas accorder une importance trop grande au phénotypes rares. Avec des données qualitatives de type présence /absence d'allèle, et lorsque la distance du chi-2 est utilisée, l'importance accordée aux allèles rares et aux combinaisons alléliques originales devient très grande.

Pour la gestion et la description de la collection *ex-situ* des sorghos cultivés, les trois sous-ensembles que nous avons constitués permettent d'avoir un ensemble global retenant une bonne part de la diversité des sorghos. De plus, la représentation des variétés insensibles, moins fréquentes dans la collection de base, est bien assurée dans le sous-ensemble L. Les variétés présentant des caractéristiques morpho-agronomiques d'intérêt, c'est à dire des morphotypes spécifiques de région particulière, sont elles aussi bien représentées dans le sous-ensemble T. L'ensemble constitué par la réunion des trois prélèvements peut ainsi prétendre à assurer une bonne représentation de la diversité des sorghos cultivés. Après caractérisation et évaluation exhaustive, cet ensemble peut constituer un point d'entrée dans la collection de base, et ainsi présenter un intérêt pour les gestionnaires de la banque de semences de sorgho. Ces derniers pourront fournir à différents utilisateurs un ensemble diversifié d'accessions pouvant servir de base à la constitution de collections de travail.

Aujourd'hui, de nombreuses collections ont été constituées dans le monde ; certaines ont atteint des tailles considérables. Est-il raisonnable d'envisager l'évaluation moléculaire de ces collections ? Quel intérêt offre une telle évaluation ?

Dans ce travail nous avons analysé près de 600 accessions à l'aide de 15 couples d'amorces permettant l'amplification spécifique de 15 locus microsatellites. Ce travail réalisé dans les conditions techniques les plus favorables a nécessité 4 mois pour le seul génotypage de ces accessions. Un travail de routine, effectué dans d'autres conditions (pas d'impératif de thèse, autre équipement de laboratoire) et concernant un nombre d'accessions du même ordre de grandeur que celui de la collection de base étudiée (par exemple 22 000) nécessiterait selon le type d'appareillage disponible entre 6 mois et 2 ans pour un jeu de 5 locus en multiplex et entre 1,5 à 5 ans pour 3 jeux de 5 locus (sans considérer les étapes d'extraction, dosage, vérification...). En comparaison, l'évaluation morpho-agronomique peut difficilement concerner plus de 1 000 accessions à la fois par an et nécessiterait donc plus de 20 ans au total. Il est donc techniquement tout à fait envisageable de procéder à l'évaluation moléculaire d'une grande collection lorsque le choix des marqueurs à utiliser est judicieux.

La diversité moléculaire (estimée à partir des 15 locus microsatellites) de chacune des core collections est forte et les indices de diversité (indices de Nei,



richesse allélique...) sont globalement du même ordre de grandeur. Cependant, des modifications dans les distributions alléliques de certains locus ont été notées entre les trois core collections L, PCS et T. De plus, la sélection d'un sous-ensemble constitué à partir des données moléculaires obtenues pour 5 locus modifie peu la diversité morpho-agronomique retenue. Les marqueurs moléculaires peuvent donc permettre un échantillonnage de la diversité morpho-agronomique de manière assez satisfaisante. Dans le cas du sorgho, 5 locus semblent suffisants pour cet échantillonnage.

Dans ce travail, nous avons choisi les marqueurs microsatellites en raison de leur simplicité d'utilisation, et du niveau de polymorphisme révélé. Chez le sorgho, ces marqueurs sont employés dans des études de diversité génétique (Djè et al., 1999), de génotypage (Dean et al., 1999) ou de cartographie du génome (Taramino et al., 1997). Le polymorphisme révélé au sein d'un ensemble d'environ 600 accessions à partir de 15 locus répartis sur l'ensemble du génome du sorgho est très important, 19 allèles en moyenne par locus, mais 206 allèles sur 289 sont rares (fréquence inférieure à 5%). Dans un prélèvement aléatoire, si le nombre d'individus portant ces allèles rares est petit, ceux-ci ont peu de chances d'être échantillonnés. Par contre, l'utilisation d'une stratégie PCS sur des données de type présence/absence incluant de très nombreux allèles rares peut biaiser la représentativité de la diversité non prise en compte dans l'échantillonnage. Nous avons vu que lorsque tous les allèles sont considérés simultanément, beaucoup de variétés bicolor sont échantillonnées au détriment des guinea sous-représentés. Par contre, lorsque seuls les allèles fréquents sont pris en compte, la représentation raciale dans l'échantillon paraît être plus équilibrée et représentative de la diversité morpho-agronomique de l'ensemble initial.

En résumé, peu de locus peuvent suffire à l'évaluation moléculaire. Aujourd'hui, ce type d'évaluation est beaucoup plus compétitif qu'une évaluation morpho-agronomique. Par contre, toutes les données ne peuvent pas être utilisées telles quelles. Une réflexion s'impose quant à l'origine et au sens du polymorphisme observé, notamment vis-à-vis des allèles rares. L'évaluation moléculaire n'offre pas d'intérêt direct car elle concerne une diversité non observable par le sélectionneur. Par contre, cette évaluation peut permettre de repérer les redondances et donc de procéder à un "nettoyage" de la collection de base (Dean et al., 1999). Elle peut aussi et doit surtout servir à définir le sous-ensemble (core collection) qui fera l'objet des différentes évaluations (morpho-agronomiques, résistances, qualité technologique des

grains etc...). Les résultats de ces investigations fourniront alors les éléments nécessaires pour réaliser des évaluations ponctuelles et ciblées au niveau de la collection de base, favorisant ainsi l'identification de types particuliers.

On ne saurait conclure ce travail sans penser au devenir des ressources génétiques dans ce 3<sup>ème</sup> millénaire. La conservation des plantes par l'intermédiaire des espèces cultivées, et/ou des espèces sauvages plus ou moins apparentées a permis et permet encore aujourd'hui de préserver une certaine forme de diversité génétique. Ce concept n'est-il pas en pleine évolution avec les possibilités offertes par la transformation génétique ?

Conservation *ex-situ*, *in-situ*, on-farm ; ces différents concepts ont été proposés pour maintenir tour à tour des génotypes isolés, puis un fonctionnement de type populations, enfin une organisation évolutive globale vis-à-vis de l'écosystème (biodiversité). Mais nous avons vu aussi qu'aucune de ces méthodes n'est aujourd'hui pleinement satisfaisante et/ou réalisable. Peut-on développer des moyens pour que ces modes de conservation puissent atteindre leurs objectifs ? Ainsi, il pourrait s'agir approfondir les connaissances sur la dynamique d'évolution des variétés traditionnelles, afin de maintenir au mieux la diversité et de bien appréhender les risques d'érosion génétique lors d'une conservation *ex-situ*. Il faudrait aussi pouvoir faire face, financièrement, à la maintenance à long terme de milliers d'accessions et au mode de conservation *in-situ*. Le développement d'une économie de marché pourrait donner les moyens aux agriculteurs d'assurer une préservation des ressources génétiques on-farm.

Et demain, que conserverons-nous ? Les banques BAC (Bacterial Artificial Chromosome) permettent de disposer et de conserver des gènes et bientôt les puzzles pour certains organismes seront reconstitués. Si on dispose aussi de la mutagenèse dirigée, est-il pas envisageable de considérer les ressources génétiques comme un ensemble de gènes multiorganismes pouvant être assemblés et rendus fonctionnels pour constituer un organisme présentant des caractéristiques sur demande ?

Les connaissances en génétique ont progressé à grand pas depuis la découverte de la structure de l'ADN de Watson et Crick en 1953, notamment en raison de l'accès à des outils technologiques de plus en plus performants. On pourra bientôt faire une lecture linéaire de l'information génétique contenue dans un génome (génomique), et une lecture "fonctionnelle" de ce génome (protéomique). Dans une autre étape, il sera

indispensable de procéder à une lecture tridimensionnelle de ce génome pour en comprendre toutes les subtilités de fonctionnement/dysfonctionnement.

Comment évoluera alors le concept des ressources génétiques ? Les avancées de la génomique auront nécessairement des répercussions sur la façon de considérer les ressources génétiques. D'une part, pour des raisons économiques, la conservation à long terme devra s'orienter sur un nombre restreint de variétés. D'autre part, la conservation des gènes mènera très certainement au développement de techniques de conservation à mettre en oeuvre pour disposer de nombreux gènes facilement intégrables.



## RÉFÉRENCES BIBLIOGRAPHIQUES

## RÉFÉRENCES BIBLIOGRAPHIQUES

- Aldrich PR, Doebley J, 1992. Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. Theor Appl Genet 85 : 293-302
- Aldrich PR, Doebley J, Schertz KF, Stec A, 1992. Patterns of allozyme variation in cultivated and wild *Sorghum bicolor*. Theor Appl Genet 85 : 451-460
- Appa Rao S, Prasada Rao KE, Mengesha MH, Gopal Reddy V, 1996. Morphological diversity in sorghum germplasm from India. Gen Res and Crop Evol 43 : 559-567
- Arasu NT, Rajanaidu N, 1975. Conservation and utilization of genetic resources in the oil palm. South East Asian Plant Genetic Resources Symposium, Bogor, pp 182-186
- Asante SA, 1995. Sorghum quality and utilization. African Crop Science Journal 3 (2) : 231-240
- Astley D, 1992. Preservation of genetic diversity and accession integrity. Field crops Research 29 : 205-224
- Ayana A, Bekele E, 1999. Multivariate analysis of morphological variation in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from Ethiopia and Eritrea. Gen Res and Crop Evol 46 : 273-284
- Balfourier F, Charmet, G, 1994. Etude méthodologique de la conservation de ressources génétiques de ray-grass anglais (graminée fourragère) par multiplication en pools de populations naturelles. Genet Sel Evol 26 (Suppl 1) : 203-218
- Balfourier F, Charmet G, Prosperi J-M, Goulard M, Monestiez P, 1998. Comparison of different spatial strategies for sampling a core collection of natural populations of fodder crops. Genet Sci Evol 30 (Suppl.1) : 215-235
- Barcelos E, 1998. Etude de la diversité génétique du genre *Elaeis* (*E.oleifera* (Kunth) Cortes et *E.guineensis* Jacq.) par marqueurs moléculaires (RFLP et AFLP). Diplôme de Doctorat, Ecole Nationale Supérieure Agronomique de Montpellier, 137 p
- Basigalup DH, Barnes, DK, Stuker RE, 1995. Development of a core collection for perennial *Medicago* plant introduction. Crop Sci 35 : 1163-1168
- Bataillon T, 1994. Comparaison de diverses stratégies d'échantillonnage pour la constitution de *core collections* de ressources génétiques végétales (études par simulation informatique). Mémoire de DEA, Institut National Agronomique, Paris, 43 p
- Bataillon TM, David JL, Schoen DJ, 1996. Neutral genetic markers and conservation genetics : simulated germplasm collections. Genetics 144 : 400-417

- Bezançon G, 1999. Le mil. In : Hamon P, Seguin M, Perrier X, Glaszmann JC, (eds) Diversité génétique des plantes tropicales cultivées. CIRAD, Montpellier, pp 309-326
- Bisht IS, Mahajan RK, Loknathan TR, Agrawal RC, 1998a. Diversity in Indian sesame collection and stratification of germplasm accessions in different diversity groups. *Gen Res and Crop Evol* 45 (4) : 325-335
- Bisht IS, Mahajan RK, Patel DP, 1998b. The use of characterisation data to establish the Indian mungbean core collection and assessment of genetic diversity. *Gen Res and Crop Evol* 45 (2) : 127-133
- Bogyo TP, Porceddu E, Perrino P, 1980. Analysis of sampling strategies for collecting genetic material. *Economic Botany* 34 (2) : 160-174
- Bouton JH, 1996. Screening the Alfalfa core collection for acid soil tolerance. *Crop Sci* 36 : 198-200
- Brown AHD, 1995. The core collection at the crossroads. In : Hodgkin T et al. (eds) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK, pp 3-19
- Brown AHD, 1989a. The case for core collections. In : Brown AHD, Frankel OH, Marshall DR, Williams JT, (eds) The use of plant genetic resources, Cambridge, pp 135-156
- Brown AHD, 1989b. Core collections : a practical approach to genetic resources management. *Genome* 31 : 818-824
- Brown AHD, Brubaker CL, Grace JP, 1997. Regeneration of germplasm samples : wild versus cultivated plant species. *Crop Sci* 37 : 7-13
- Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzales-Candelas F, Kresovich S, 1996. Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. *Theor Appl Genet* 93 : 190-198
- Brush SB, 1995. In situ conservation of landraces in centers of crop diversity. *Crop Sci* 35 : 346-354
- Carlier J, Lebrun, M.H., Zapater, M.F., Dubois, C., Mourichon, X., 1996. Genetic structure of the global population of banana black leaf streak fungus, *Mycosphaerella fijiensis*. *Molecular Ecology* 5 : 499-510
- Casler M, 1995. Patterns of variation in a collection of perennial ryegrass accessions. *Crop Sci* 35 : 1169-1177
- CATIE, 1999. <http://www.catie.ac.cr/>
- CGIAR, 1999. <http://www.noc1.cgiar.org/>



- Chantereau J, Arnaud M, Ollitrault P, Nabayaogo P, Noyer J-L, 1989. Etude de la diversité morphophysologique et classification des sorghos cultivés. *L'agronomie tropicale* 44 (3) : 223-231
- Chantereau J, Kondombo C, 1994. Estimation du taux d'allogamie chez les sorghos de la race guinea. In : Menyonga JM, Bezuneh T, Yayock JY, Soumana I, (eds) *Progress in food grain research and production in semi-arid Africa*. Safgrad Inter-Network Conference, Niamey, Niger, 03-07 avril 1991, pp 309-313
- Chantereau J, Luce C, Hamada MAG, Trouche G, 1998. L'utilisation des sorghos guinea en sélection dans le programme conjoint Icrisat-Cirad sur le sorgho. In : Ratnadass A, Chantereau J, Gigou J, (eds) *Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre*. Actes de l'atelier de restitution du programme conjoint sur le sorgho ICRISAT-CIRAD, Bamako, Mali, pp 65-72
- Chapman CGD, 1989. Collections strategies for the wild relatives of field crops. In : Brown AHD, Frankel OH, Marshall DR, Williams JT, (eds) *The use of plant genetic resources*. Cambridge, pp 263-279
- Charmet G, Balfourier F, Monestiez P, 1994. Hierarchical clustering of perennial ryegrass populations with geographic continuity constraint. *Theor Appl Genet* 88 : 42-48
- Charmet G, Balfourier F, 1995. The use of geostatistics for sampling a core collection of perennial ryegrass populations. *Gen Res and Crop Evol* 42 : 303-309
- Collectif, 1996. *Lost Crops of Africa*. National academy press, Washington, DC, 383 p
- Cordeiro CMT, Morales EAV, Ferreira P, Rocha DMS, Costa IRS, Valois ACC, Silva S, 1995. Towards a Brazilian core collection of cassava. In : Hodgkin T, Brown AHD, van Hintum TJL, Morales EAV, (eds) *Core collections of plant genetic resources*. John Wiley & Sons, Chichester, UK, pp 155-168
- Crossa J, DeLacy IH, Taba S, 1995. The use of multivariate methods in developing a core collection. In : Hodgkin T, Brown AHD, van Hintum TJL, Morales EAV, (eds) *Core collections of plant genetic resources*. John Wiley & Sons, Chichester, UK, pp 77-92
- Dahlberg JA, Spinks MS, 1995. Current status of the US sorghum germplasm collection. *ISMN* 36 : 4-12
- de Wet MJM, 1978. Systematics and evolution of sorghum sect. sorghum (Gramineae). *Amer J Bot* 65 (4) : 477-484
- de Wet MJM, Huckabay JP, 1967. The origin of *Sorghum bicolor*. II Distribution and domestication. *Evolution* 21 : 787-802
- Dean RE, Dahlberg JA, Hopkins MS, Mitchell SE, Kresovich S, 1999. Genetic redundancy and diversity among 'Orange' accessions in the U.S. national sorghum collection as assessed with simple sequence repeat (SSR) markers. *Crop Sci* 39 : 1215-1221

- Deu M, Gonzales-de-Leon D, Glaszman JC, Degremont I, Chanterreau J, Lanaud C, Hamon P, 1994. RFLP diversity in cultivated sorghum in relation to racial differentiation. *Theor Appl Genet* 88 : 838-844
- Deu M, Hamon P, Chanterreau J, Dufour P, D'Hont A, Lanaud C, 1995. Mitochondrial DNA diversity in wild and cultivated sorghum. *Genome* 38 : 635-645
- Deu M, Hamon P, Bonnot F, Chanterreau J, 1999. Le Sorgho. In : Hamon P, Seguin M, Perrier X, Glaszmann JC, (eds) *Diversité génétique des plantes tropicales cultivées*, CIRAD, Montpellier, France, pp 351-382
- Divaret I, Margale E, Thomas G, 1999. RAPD markers on sweet bulks efficiently assess the genetic diversity of a *Brassica oleracea* L. collection. *Theor Appl Genet* 98 : 1029-1035
- Diwan N, Bauchan GR, McIntosh MS, 1994. A core collection for the United States annual *Medicago* germplasm collection. *Crop Sci* 34 : 279-285
- Diwan N, McIntosh, MS, Bauchan GR, 1995. Methods of developing a core collection of annual *Medicago* species. *Theor Appl Genet* 90 : 755-761
- Djè Y, Ater M, Lefèbvre C, Vekemans X, 1998. Patterns of morphological and allozyme variation in sorghum landraces of Northwestern Morocco. *Gen Res and Crop Evol* 45 : 541-548
- Djè Y, Forcioli D, Ater M, Lefèbre C, Vekemans X, 1999. Assessing population genetic structure of sorghum landraces from North-western Morocco using allozyme and microsatellite markers. *Theor Appl Genet* 99 (157-163) :
- Doggett H, 1988. Sorghum. 2<sup>nd</sup> edition Longman Scientific & Technical, UK copublished in the United States with John Wiley & Sons, Inc., New York, 512 p
- Doggett H, Prasada Rao KE, 1995. Sorghum. In : Simmonds Edit, Longman, London, 496 p
- Dussert S, Chabrillange N, Anthony F, Engelmann F, Recalt C, Hamon S, 1997. Variability in storage response within a coffee (*Coffea* ssp.) core collection under slow growth conditions. *Plant Cell Reports* 16 : 344-348
- Eberhart SA, Bramel-Cox PJ, Prasada Rao KE, 1998. Preserving genetic resources. In *Proceedings of the International Conference on Genetic Improvement of Sorghum and Pearl Millet*, Lobbock, Texas 23-27 Sept. 1996, pp 25-41
- Ellis PR, Pink DAC, Phelps K, Jukes PL, Breeds SE, Pinnegar AE, 1998. Evaluation of a core collection of *Brassica oleracea* accessions for resistance to *Brevicoryne brassicae*, the cabbage aphid. *Euphytica* 103 : 149-160
- Engels JMM, Ramanatha Rao R, 1995. Regeneration of seed crops and their wild relatives. ICRISAT, Hyderabad, India, 167 p
- FAO, 1999. <http://apps.fao.org/>

- Fliedel G, Marti A, Thiebaut S, 1996. Caractérisation et valorisation du sorgho. Les bibliographies du CIRAD. CIRAD-CA, (eds), Montpellier, 404 p
- Ford-Lloyd B, Jackson M, 1986. Plant genetic resources : an introduction to their conservation and use. Edward Arnold, London, 146 p
- Galwey NW, 1995. Verifying and validating the representativeness of a core collection. In : Hodgkin T, Brown AHD, van Hintum TJL, Morales EAV, (eds) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK, pp 187-198
- Hamon S, van Sloten DH, 1989. Caractérisation and evaluation of Okra. In : Brown AHD, Frankel OH, Marshall DR, Williams JT, (eds) The use of plant genetic resources. Cambridge, pp 173-196
- Hamon S, Noirot M, Anthony F, 1995. Developing a coffee core collection using principal components score strategy with quantitative data. In : Hodgkin T, Brown AHD, van Hintum TJL, Morales EAV, (eds) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK, pp 117-126
- Hamon S, Dussert S, Deu M, Hamon P, Seguin M, Glaszmann J-C, Grivet L, Chantreau J, Chevallier M-H, Flori A, Lashermes P, Legnate H, Noirot M, 1998. Effects of quantitative and qualitative principal component score strategies on the structure of coffee, rubber tree, rice, and sorghum core collections. Genet Sel Evol 30 (Suppl.1) : 237-258
- Hamrick JL, Godt MJW, 1995. Conservation genetics of endemic plant species. In : Avise JC, Hamrick JL, (eds) Conservation genetics. Case histories from nature. Chapman&Hall, New York, pp 281-304
- Harlan JR, de Wet MJM, 1971. Toward a relational classification of cultivated plants. Taxon 20 (4) : 509-517
- Harlan JR, de Wet MJM, 1972. A simplified classification of cultivated sorghum. Crop Sci 12 : 172-176
- Harlan JR, Stemler A, 1976. The races of sorghum in Africa. In : Harlan JR, de Wet MJM, Stemler ABL, (eds) Origins of African Plant Domestication. Mouton publishers, The Hague-Paris, pp 465-478
- Hintum ThJL van, Haalman D, 1994. Pedigree analysis for composing a core collection of modern cultivars, with examples from barley (*Hordeum vulgare* s. lat.). Theor Appl Genet 88 : 70-74
- Hintum ThJL van, 1995. Hierarchical approaches to the analysis of genetics diversity in crop plants. In : Sons JW, (eds) Core collections of plant genetic resources. Chichester, UK, pp23-24
- Hintum ThJL van, Von Bothmer R, Visser DL, 1995. Sampling strategies for composing a core collection of cultivated barley (*Hordeum vulgare* s.lat.) collected in China. Hereditas 122 : 7-17



- Hodgkin T, Brown AHD, van Hintum ThJL, Morales EAV, 1995. Future directions. In : Hodgkin T, Brown AHD, van Hintum TJL, Morales EAV, (eds) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK, pp 253-260
- Holbrook CC, Anderson WF, Pittman RN, 1993. Selection of a core collection from the U.S. germplasm collection of peanut. *Crop Sci* 33 : 859-861
- Holbrook CC, Anderson, WF, 1995. Evaluation of a core collection to identify resistance to late leafspot in peanut. *Crop Science* 35 : 1700-1702
- Holcomb J, Tolbert DM, Jain SK, 1977. A diversity analysis of genetic resources in rice. *Euphytica* 26 : 441-450
- Huamán Z, Aguilar C, Ortiz R, 1999. Selecting a Peruvian sweetpotato core collection on the basis of morphological, eco-geographical, and disease and pest reaction data. *Theor Appl Genet* 98 : 840-844
- Hutcheson K, 1970. A test for comparing diversities based on the Shannon formula. *J Theor Biol* 29 : 151-154
- IBPGR ICRISAT, 1993. Descriptors for sorghum [*Sorghum bicolor* (L) Moench]. ICRISAT, Patancheru, India, 176 p
- ICRISAT, 1996. The world sorghum and millet economies. Facts, trends and outlook. ICRISAT & FAO, 68 p
- Igartua E, Gracia, MP, Lasa JM, Medina B, Molina-Cano JL, Montoya JL, Romagosa I, 1998. The Spanish barley core collection. *Gen Res and Crop Evol* 45 : 475-481
- IPGRI, 1995. Directory of European PGR collections - Holdings. Frison EA, Serwinski J, (eds) IPGRI, Rome, 499 p
- Jacquemard J-C, Baudouin L, Noiret J-M, 1997. Le plamier a huile. In : Charrier A, Jacquot M, Hamon S, Nicolas D, (eds) L'amélioration des plantes tropicales, CIRAD, pp 507-532
- Jain SK, Qualset CO, Bhatt GM, Wu KK, 1974. Geographic patterns of phenotypic diversity in a world collection of durum wheats. *Crop Sci* 15 : 700-704
- Jarvis DI, 1999. Strengthening the scientific basis of *in situ* conservation of agricultural biodiversity on-farm. *Botanica Lithuanica* 2 : 79-90
- Jung HG, Sheaffer CC, Barnes DK, Halgerson JL, 1997. Forage quality variation in the U.S. alfalfa core collection. *Crop Sci* 37 : 1361-1366
- Lanaud C, Motamayor J-C, Sounigo O, 1999. Le cacaoyer. In : Hamon P, Seguin M, Perrier X, Glaszmann JC, (eds) Diversité génétique des plantes tropicales cultivées, CIRAD, Montpellier, pp 141-174

- Lerceteau E, Robert T, Petiard V, Crouzillat D, 1997. Evaluation of the extent of genetic variability among *Thebroma cacao* accessions using RAPD and RFLP markers. *Theor Appl Genet* 95 : 10-19
- Li Y, Wu S, Cao YS, Zhang X, 1996. A phenotypic diversity analysis of foxtail millet (*Setaria italica* (L.) P.Beauv.) landraces of Chinese origin. *Gen Res and Crop Evol* 43 : 377-384
- Louette D, 1994. Gestion traditionnelle des variétés de maïs dans la réserve de la Biosphère Sierra de Manantlan (RBSM, états de Jalisco et Colima, Mexique) et conservation *in situ* des ressources génétiques de plantes cultivées. Thèse de doctorat, Ecole Nationale Supérieure Agronomique de Montpellier, 245 p
- Mackay MC, 1995. One or many core collections? In : Hodgkin T, Brown AHD, van Hintum TJL, Morales EAV, (eds) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK, pp 199-210
- Mahajan RK, Bisht IS, Agrawal RC, Rana RS, 1996a. Studies on South Asian okra collection : methodology for establishing a representative core set using characterization data. *Gen Res and Crop Evol* 43 : 249-255
- Mahajan RK, Kochhar S, Agrawal RC, Chandel KPS, 1996b. Developing a core set or subsets of large germplasm collections : an overview. *Indian J Pl Genet Resources* 9 (1) : 11-30
- Mann JA, Kimber CT, Miller FR, 1983. The origin and early cultivation of Sorghums in Africa. College Station, TX, USA, 21 p
- Marshall DR, 1989. Limitations to the use of germplasm collections. In Brown AHD, Frankel OH, Marshall DR, Williams JT, (eds), The use of plant genetic resources. Cambridge, pp 105-122
- Mengesha MH, Prasada Rao KE, 1982. Current situation and future of sorghum germplasm. In : House LR, Mughogho LK, Peacock JM, Mertin JV (eds), Sorghum in the Eighties : Proceedings of the international symposium on sorghum, ICRISAT, Patancheru, A.P., India, pp 323-334
- Menkir A, Goldsbrough P, Ejeta G, 1997. RAPD based assessment of genetic diversity in cultivated races of Sorghum. *Crop Sci* 37 : 564-569
- Morden CW, Doebley J, Schertz KF, 1990. Allozyme variation among the spontaneous species of *Sorghum* section Sorghum (Poaceae). *Theor Appl Genet* 80 : 296-304
- Morden CW, Doebley JF, Schertz KF, 1989. Allozyme variation in old world races of *Sorghum bicolor* (Poaceae). *Amer J Bot* 76 (2) : 247-255
- Murty BR, Govil JN, 1967. Description of 70 groups in genus sorghum based on a modified Snowden's classification. *The Indian Journal of Genetics & Plant Breeding* 27(Special Number) : 75-91

- Nei N, 1973. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70 (12) : 3321-3323
- Noirot M, Hamon S, Anthony F, 1996. The principal component scoring : a new method of constituting a core collection using quantitative data. *Gen Res and Crop Evol* 43 : 1-6
- O'Brien S, 1994. A role for molecular genetics in biological conservation. *Proc Natl Acad Sci USA* 91 : 5748-5755
- Ollitrault P, 1987. Evaluation génétique des sorghos cultivés (*Sorghum bicolor* L. Moench) par l'analyse conjointe des diversités enzymatique et morphophysiologique; relation avec les sorghos sauvages. Thèse de doctorat, Paris-XII, Orsay, 187 p
- Ollitrault P, Arnaud M, Chantereau J, 1989. Polymorphisme enzymatique des sorghos. II-Organisation génétique et évolutive des sorghos cultivés. *L'agronomie tropicale* 44 (3) : 211-221
- Ollitrault P, Noyer J-L, 1989. Polymorphisme enzymatique des sorghos. III-Identification et classification de variétés améliorées de l'IRAT. *L'agronomie tropicale* 45 (1) : 59-66
- Ortiz R, Ruiz-Tapia EN, Mujica-Sanchez A, 1998. Sampling strategy for a core collection of Peruvian quinoa germplasm. *Theor Appl Genet* 96 : 475-483
- Ortiz R, Madsen S, Ruiz-Tapia EN, Jacobsen S-E, Mujica-Sanchez A, Christiansen JL, Stølen O, 1999. Validating a core collection of Peruvian quinoa germplasm. *Gen Res and Crop Evol* 46 : 285-290
- Pedersen JF, Toy JJ, Johnson B, 1998. Natural outcrossing of Sorghum and Sundangrass in the central great plains. *Crop Sci* 38 : 937-939
- Pederson GA, Fairbrother TE, Greene SL, 1996. Cyanogenesis and climatic relationships in U.S. white clover germplasm collection and core subset. *Crop Sci* 36 : 427-433
- Peeters JP, Matrinelli JA, 1989. Hierarchical cluster analysis as a tool to manage variation in germplasm collections. *Theor. Appl. Genet.* 78 : 42-48
- Pernès J, 1984. Gestion des ressources génétiques des plantes. Tome 2 : Manuel. Agence de coopération culturelle et technique, Paris 346 p
- Perry MC, McIntosh MS, 1991. Geographical patterns of variation in the USDA soybean collection : I. Morphological traits. *Crop Sci* 31 : 1350-1355
- Polignano GB, Alba E, Ugenti P, Scippa G, 1999. Geographical patterns of variation in Bari faba bean germplasm collection. *Gen Res and Crop Evol* 46 : 183-192
- Prasada Rao KE, Gopal Reddy VG, 1989. Status of the world collection of sorghum at ICRISAT. ICRISAT, Patancheru



- Prasada Rao KE, Mengesha MH, Reddy VG, 1989. International use of sorghum germplasm collection. In : Brown ADH, Frankel OH, Marshall DR, Williams JT, (eds) The use of plant genetic resources. Cambridge, pp 49-67
- Prasada Rao KE, Gopal Reddy VG, Stenhouse JW, 1995. Sorghum genetic resources at ICRISAT Asia Center. ISMN 36 : 15-18
- Prasada Rao KE, Ramanatha Rao V, 1995b. The use of characterisation data in developing a core collection of sorghum. In : Hodgkin T, Brown AHD, van Hintum TJL, Morales EAV, (eds) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK, pp 109-116
- Quinby, J.R. 1966. Fourth maturity locus in sorghum. Crop Sci 6 : 516-518
- Quinby, J.R. 1974. Sorghum improvement and the genetics of growth. Texas University, Texas A&M University Press, 108 p
- Rami J-F, Dufour P, Trouche G, Fliedel G, Mestres C, Davrieux F, Blanchard P, Hamon P, 1998. Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench). Theor Appl Genet 97 : 605-616
- Raymond M, Rousset F, 1995. Population genetics software for exact tests and eucmenicism. Heredity 86 : 248-249
- Rincon F, Johnson B, Crossa J, Taba S, 1997. Identifying subsets of maize accessions by three-mode principal component analysis. Crop Sci 37 : 1936-1943
- Rooney, L.W., & F.R. Miller, 1982. Variation in the structure and kernel characteristics of sorghum. Proceedings of the International symposium in sorghum grain quality. ICRISAT, Patancheru, A.P., India, pp 143-162
- Schoen DJ, Brown AHD, 1993. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc Natl Acad Sci USA 90 : 1063-10627
- Schoen DJ, Brown AHD, 1995. Maximising allelic diversity in core collections of wild crop relatives : the role of genetic markers. In : Sons JW, (eds) Core collections of plant genetic resources. Chichester, UK, pp 55-76
- Skinner DZ, Bauchan, GR, Auricht G, Hughes S, 1999. A method for the efficient mangement and utilization of large germplasm collections. Crop Sci 39 : 1237-1242
- Skroch PW, Nienhuis J, Beebe S, Tohme J, Pedraza F, 1998. Comparison of Mexican common bean (*Phaseolus vulgaris* L.) core and reserve germplasm collections. Crop Sci 38 : 488-496
- Snowden JD, 1936. The cultivated races of *Sorghum*. London, Adlard, 274 p
- Soleri D, Cleaveland D, 1993. Seeds of strength for Hopis and Zunis. Seedling 10 (4). In <http://www.ciesin.org/>

- Spagnoletti Zeuli PL, Qualset CO, 1993. Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. *Theor Appl Genet* 87 : 295-304
- Spagnoletti Zeuli PL, Qualset CO, 1987. Geographical diversity for quantitative spike characters in world collection of durum wheat. *Crop Sci.* 27 : 235-241
- StatSoft, 1995. Statistica for Windows. USA, StatSoft, Tulsa, OK, USA
- Stemler, A.B.L., J.R. Harlan & J.M.J. de Wet, 1975. Caudatum sorghums and speakers of chari-nile languages in Africa. *Journal of African History* XVI (2) : 161-183
- Stenhouse JW, Prasada Rao KE, Gopal Reddy VG, Appa Rao S, 1997. Sorghum. In : Fucillo D, Sears L, Stapleton P, (eds) *Biodiversity in trust*. Cambridge University Press, Cambridge, UK, pp 292-308
- Taba S, Diaz J, Franco J, Crossa J, 1998. Evaluation of Caribbean maize accessions to develop a core subset. *Crop Sci* 38 : 1378-1386
- Tao Y, Manners JM, Ludlow MM, Henzell RG, 1993. DNA polymorphisms in grain sorghum (*Sorghum bicolor* (L.) Moench). *Theor Appl Genet* 86 : 679-688
- Taramino G, Tarchini R, Ferrario S, Lee M, Pe ME, 1997. Characterization and mapping of single sequence repeats (SSRs) in *Sorghum bicolor*. *Theor Appl Genet* 95 : 66-72
- Teshome A, Baum BR, Fahrig L, Torrance JK, Arnason TJ, Lambert JD, 1997. Sorghum [*Sorghum bicolor* (L.) Moench] landrace variation and classification in North Shewa and South Welo, Ethiopia. *Euphytica* 97 : 255-263
- Tohme J, Jones P, Beebe S, Iwanaga M, 1995. The combined use of agroecological and characterisation data to establish the CIAT *Phaseolus vulgaris* core collection. In : Hodgkin T, Brown AHD, van Hintum TJL, Morales EAV, (eds) *Core collections of plant genetic resources*. John Wiley & Sons, Chichester, UK, pp 95-107
- Tolbert DM, Qualset CO, Jain SK, Craddock JC, 1978. A diversity analysis of a world collection of barley. *Crop Sci* 19 : 789-794
- Trouche G, Vaksman M, Chanterreau J, Kouressy M, Maiga HD, Barro C, 1998. Etude du déterminisme génétique du photopériodisme des sorghos guinea. In : Ratnadass A, Chanterreau J, Gigou J, (eds) *Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre*. Actes de l'atelier de restitution du programme conjoint sur le sorgho ICRISAT-CIRAD, Bamako, Mali, pp 37-47
- Vaughan DA, Jackson MT, 1995. The core as a guide to the whole collection. In : Hodgkin T, Brown AHD, van Hintum TJL, Morales EAV, (eds) *Core collections of plant genetic resources*. John Wiley & Sons, Chichester, UK, pp 229-239

- Weltzien E, 1989. Differentiation among barley landrace populations from the Near East. *Euphytica* 43 : 29-39
- Wendorf F, Close AE, Schild R, Wasylikowa K, Housley RA, Harlan JR, Krolik H, 1992. Saharan exploitation of plants 8,000 years BP. *Nature* 359 : 721-724
- Wu KK, Jain SK, 1977. A note on germ plasm diversity in the world collections of safflower. *Economic Botany* 31 : 72-75
- Yonezawa K, Nomura T, Morishima H, 1995. Sampling strategies for use in stratified germplasm collections. In : Hodgkin T, Brown AHD, van Hintum TJJ, Morales EAV, (eds) *Core collections of plant genetic resources*. John Wiley & Sons, Chichester, UK, pp 35-53



## ANNEXES

**Annexe 1** : indices de diversité de Shannon-Weaver (valeurs standardisée  $\pm$  variance) calculés en fonction des races (les 5 races principales et les deux races intermédiaires les plus fréquentes) et des origines géographiques (les 5 régions géographiques les plus représentées dans la collection). L'effectif de chaque groupe est donné entre parenthèses.

race	Afrique orientale	Afrique centrale	sous-continent indien	Afrique australe	Afrique occidentale	total
bicolor	0,64 $\pm$ 0,00 (179)	0,64 $\pm$ 0,01 (50)	0,64 $\pm$ 0,00 (173)	0,53 $\pm$ 0,01 (34)	0,71 $\pm$ 0,00 (85)	0,68 $\pm$ 0,00 (521)
caudatum	0,67 $\pm$ 0,00 (2431)	0,64 $\pm$ 0,00 (546)	0,60 $\pm$ 0,00 (84)	0,54 $\pm$ 0,00 (131)	0,59 $\pm$ 0,00 (1495)	0,67 $\pm$ 0,00 (4687)
durra	0,57 $\pm$ 0,00 (1442)	0,59 $\pm$ 0,00 (76)	0,51 $\pm$ 0,00 (2442)	0,49 $\pm$ 0,01 (55)	0,55 $\pm$ 0,00 (456)	0,56 $\pm$ 0,00 (4471)
guinea	0,70 $\pm$ 0,01 (56)	0,61 $\pm$ 0,00 (801)	0,55 $\pm$ 0,00 (560)	0,50 $\pm$ 0,01 (44)	0,58 $\pm$ 0,00 (2405)	0,62 $\pm$ 0,00 (3866)
kafir	0,36 $\pm$ 0,04 (7)	0,47 $\pm$ 0,01 (61)	0,15 $\pm$ 0,16 (3)	0,44 $\pm$ 0,00 (429)	0,51 $\pm$ 0,17 (3)	0,46 $\pm$ 0,00 (503)
DC	0,64 $\pm$ 0,00 (1658)	0,67 $\pm$ 0,00 (104)	0,65 $\pm$ 0,00 (191)	0,51 $\pm$ 0,01 (45)	0,56 $\pm$ 0,00 (609)	0,67 $\pm$ 0,00 (2607)
GC	0,74 $\pm$ 0,00 (808)	0,67 $\pm$ 0,00 (670)	0,72 $\pm$ 0,00 (95)	0,48 $\pm$ 0,00 (151)	0,68 $\pm$ 0,00 (561)	0,72 $\pm$ 0,00 (2285)
total	0,72 $\pm$ 0,00 (7601)	0,74 $\pm$ 0,00 (2438)	0,65 $\pm$ 0,00 (3866)	0,54 $\pm$ 0,00 (971)	0,70 $\pm$ 0,00 (5903)	0,74 $\pm$ 0,00 (20779)

**Annexe 2** : comparaison par un test-t entre les indices de Shannon-Weaver des différents groupes (est indiquée la valeur p), **A**) groupes raciaux, **B**) groupes géographiques

**A**

	bicolor	caudatum	durra	guinea	kafir	CB	DB	DC	GC	Total
bicolor		0,292	0,088	0,325	<0,001	0,081	0,094	0,097	0,001	<0,001
caudatum			<0,001	0,011	<0,001	0,112	0,117	0,099	<0,001	<0,001
durra				0,048	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
guinea					<0,001	0,002	<0,001	<0,001	<0,001	<0,001
kafir						<0,001	<0,001	<0,001	<0,001	<0,001
CB							0,376	0,345	0,093	0,001
DB								0,391	0,011	<0,001
DC									0,001	<0,001
GC										0,002

**B**

	Afrique orientale	Afrique occidentale	s-c indien	Afrique centrale	Afrique australe	Total
Afrique orientale		0,074	<0,001	0,039	<0,001	<0,001
Afrique occidentale			0,001	0,002	<0,001	<0,001
s-c indien				<0,001	<0,001	<0,001
Afrique centrale					<0,001	0,242
Afrique australe						<0,001



## PUBLICATIONS

Description of the Diversity of the Landrace Collection of Sorghum  
(*Sorghum bicolor* (L.) Moench) conserved at ICRISAT.

Grenier C, Deu M, Chantereau J, Reddy VG, Bramel-Cox PJ, Hamon P

*Soumis à Euphytica*

---

DESCRIPTION OF THE DIVERSITY OF THE LANDRACE COLLECTION OF  
SORGHUM (*Sorghum bicolor* (L.) Moench) CONSERVED AT ICRISAT

C. Grenier, M. Deu, J. Chantereau, V.G. Reddy, P.J. Bramel-Cox & P. Hamon

Cécile Grenier, Monique Deu and J. Chantereau CIRAD, BP 5035, 34032 Montpellier  
Cedex 1, France; V.G. Reddy and Paula J. Bramel-Cox, ICRISAT-GREP Patancheru, 502  
324 A.P., India; Perla Hamon, Université Montpellier III/IRD, route de Mende, 34199  
Montpellier Cedex 5, France

ABSTRACT

Since the seventies, ICRISAT has been responsible for the acquisition and conservation of a large sorghum collection (*Sorghum bicolor* (L.) Moench) of 36,719 accessions. Among them, 22,473 landraces have complete characterization data available from the ICRISAT sorghum database. This landrace collection was analyzed for the morpho-agronomic diversity encompassed in ten geographic regions and 15 sorghum races. Our results show that the African and Indian subcontinent are both well represented in the landrace collection. Despite its recognition as a secondary center for diversity, China is under-represented. The main cultivated races of sorghum are represented in relative frequencies in accordance with their importance in cultivation. The three intermediate forms (GC, CB, and DC) and two basic races (bicolor and caudatum) have the highest levels of phenotypic diversity. Results are discussed to highlight the main characteristics as well as peculiarities of the main races and regions. They give guidelines to breeders for a wider use of the genetic resources available in the genebank.

*Key words:* sorghum germplasm - Shannon-Weaver diversity index - morpho-agronomic variability - racial classification - geographical distribution



## INTRODUCTION

An ex situ collection for sorghum, *Sorghum bicolor* (L.) Moench, was initiated in the 1960<sup>ies</sup> by the Rockefeller Foundation as part of the Indian Agricultural Research Programs (House, 1985). This collection was transferred to the International Crops Research Institute for the Semi-Arid Tropic (ICRISAT) in Patancheru, India in 1974 (Stenhouse et al., 1997). Two decades ago, the danger of genetic erosion in traditional landraces with the release of new varieties and hybrids increased the collection and conservation effort at ICRISAT and other genebanks (Mengesha & Prasada Rao, 1982). At present, the ICRISAT collection contains 36,719 accessions acquired from 90 countries. Many of these accessions were also conserved at Fort Collins (USA) and Mayagüez (Puerto Rico, USA). In both collections, ICRISAT and USA, characterization and evaluations were undertaken. Resulting databases are available on line (<http://www.ars-grin.gov>, Dahlberg and Spinks, 1995 and <http://noc1.cgiar.org/sear-type.htm>).

Enhanced evaluation and utilization of the genetic resources held in germplasm collections requires detailed understanding of the diversity and distribution of the accessions held. Appa Rao et. al. (1996) reported on the morphological diversity of the sorghums acquired from India and conserved at ICRISAT. Ayana and Bekele (1999) have reported on the morphological variation in the national Ethiopian and Eritrean sorghum collection. However, no description of the phenotypic or genotypic diversity of the entire landrace collection have been reported for ICRISAT. Thus, the objective of this study was to assess the representation and phenotypic diversity of the sorghum landrace collection held by ICRISAT. This collection of traditional landraces is described on the basis of the geographical origin and the racial classification. The analysis of diversity included both quantitative and qualitative morpho-agronomic characters.

## MATERIAL AND METHODS

Within the ICRISAT collection only the landrace collection was considered for this analysis. This landrace collection used for this study consisted of 22,473 accessions acquired from 76 countries in the latitude range of 40° North to 40° South. Ten geographical regions were defined for this study: Eastern Africa, Western Africa, Indian Sub-continent, Central Africa, Southern Africa, Eastern Asia and Oceanic, Northern America, Southern America, Arabia, and Northern Africa and Southern Europe (Table 1). Sorghum landraces are classified into the 15 races based on the spikelets' structure according to Harlan and de Wet (1972; Table 2).

For the characterization, the accessions were grown on vertisol soils at Patancheru, India (17°25'N latitude and 78°E longitude). Field observations were made over several years from 1975 to 1996 in both the kharif (rainy) and the rabi (post-rainy) seasons. These two different environmental conditions are typical of the semi-arid tropics. Kharif is the rainy season with longer day duration (accessions sown in June and harvested in December) while rabi is the dry season with short day lengths (sown in September and harvest in April). Days to 50% flowering were recorded as the number of days from mean emergence date to the date when 50% plants had started flowering (i.e. ca. 30 plants per row per accession). Plant height was observed at maturity and recorded on ten selected plants as the length of the main stalk (cm). Both of these characters were measured in the kharif season. All following characters were recorded during the rabi season. Peduncle exertion (cm) was measured on 10 plants in the field. After harvest, one hundred seed

weight (g) was measured. Eight qualitative characters also were considered. Plant and midrib pigmentation were recorded in the field while shape of the panicle, glume and grain color, seed covering, endosperm texture, and presence/absence of a testa were recorded after harvest.

Basic statistics such as the frequency distribution, the range, the mean and the variance were calculated for the quantitative morpho-agronomic characters of the landrace collection. The distribution of landraces within the different modalities of the qualitative character, racial classification and regions of origin was also calculated. An index of representation was calculated as the ratio between the frequency of a particular race or region for a specific phenotype and the frequency of this race or region in the landrace collection. Significance of differences were tested with Chi-square tests.

Diversity within the race and the region was assessed with the Shannon-Weaver Diversity Index calculated according to Poole (1974) (in Li et al., 1996). This phenotypic diversity index was calculated for the eight qualitative characters and the four quantitative characters after transformation into four phenotypic classes as defined by the four quartiles of the landrace collection. The Shannon-Weaver Diversity Index was estimated using formula (1):

$$H_C' = - \sum_{i=1}^n p_i \log_e p_i \quad (1),$$

where for a given character C,  $n$  is the number of phenotypic classes (for the qualitative characters  $n = 2$  to 7 descriptor states, and for quantitative ones  $n = 4$  frequency classes), and  $p_i$  is the proportion of the total number of entries in the  $i^{\text{th}}$  class. Due to its additive property, Shannon-Weaver diversity indices obtained for each character were pooled for each race and region. Then, means and standard errors were calculated from the non-standardized indices, and the variance (2) was approximated by Hutcheson (1970):

$$\text{Var}(H') = [\sum p_i \log_e^2 p_i - (\sum p_i \log_e p_i)^2 / N + (n-1)/2N^2] \quad (2),$$

where  $N$  is the number of observations. The  $H'$  indices were compared by a t-test :

$$T = (H'_1 - H'_2) / [\text{Var}(H'_1) + \text{Var}(H'_2)]^{1/2},$$

with degrees of freedom : d.f. =  $[\text{Var}(H_1) + \text{Var}(H_2)]^2 / [\text{Var}(H_1)^2 / N_1 + \text{Var}(H_2)^2 / N_2]$ , where  $N_1$  and  $N_2$  are the numbers of entries in the two groups.

The indices are standardized as in formula (3) to keep each value in a range from 0 – 1 in order to estimate the importance of the phenotypic diversity.

$$\text{SDI}_C = H_C' / \log_e n \quad (3).$$

## RESULTS

The distribution of the landraces was assessed for the region and the specific country (Table 1). Three of the geographical regions are highly represented in the ICRISAT landrace collection. Eastern Africa (seven countries) was the most important. It represented 34.3% of the landrace collection. The highest percentage of these landraces were from Yemen (24.7%), Ethiopia (23.8%), and Sudan (23.7%). Egypt is poorly represented with only 12 accessions. Western Africa was represented by 14 countries in 27% of the landrace collection. Western African landraces are mainly from Cameroon (37.8%), Nigeria (18.5%) and at a lesser extend from Mali (10.4%) and Burkina Faso (8.8%). Among the countries that represent less than 5% of the considered region, only Togo and Senegal have more than 200 accessions. Only one accession originated from Cape Verde Island and one from Ivory



Coast. The third most important region is the Indian subcontinent with 17.6% of the landrace collection. In the Indian Sub-continent, 98.5% of the landraces are from India. Central African landraces came from 12 countries that represented 11.4% of the landrace collection. Among Central African landraces, 38.1% of them were from Zimbabwe. Southern Africa contains accessions from four countries and represents 5.3% of the landrace collection. South Africa Republic contributes 47.7% of the accessions. The remaining five regions represented less than 5% of the landrace collection. The Eastern Asia and Oceanic region accounts for 2%, North America for 1.7% and the remaining ones for less than 1%.

The racial classification indicated that the distribution among the 15 races was not equal (Table 2). Three main races (caudatum, durra, and guinea) and two intermediate forms (durra-caudatum (DC) and guinea-caudatum (GC)) represent the largest portion of the landrace collection (82%). Six intermediate races represented less than 1% of the landrace collection. Each race (main or intermediate) with a frequency higher than 2% in the landrace collection was found for all geographical regions. For example, caudatum, CB, DB, DC and GC were mainly acquired from Eastern Africa. The durra race was mainly acquired from the Indian subcontinent while the guinea race was from Western Africa and the kafir race from Southern Africa. The bicolor race seems to be equally represented from Eastern Africa and the Indian subcontinent.

The regional geographical representation of the landrace collection was assessed from the frequency distributions of the four quantitative and eight qualitative characters. When only regions corresponding to more than 1% of the landrace collection were considered, a comparison of the frequency of accessions from a particular region with a specific phenotype and the frequency of this region in the landrace collection based upon an index of representation (Table 3). The results presented in this analysis will be described separately for each character;

#### Days to flowering

Days to flowering had a bimodal distribution with a range from 35 to 199 days. Two main groups were distinguished; accessions that flower around 80-90 days and the late landraces that flowered in 130-140 days. For the purpose of this study the earliest material is considered as the material that flowers before 70 days and the latest landraces those that flower after 130 days. More than half the early sorghum is acquired from Eastern Africa and Southern Africa. For this later, as for Eastern Asia and Oceanic, the frequency is more than four times higher in the early landraces than in the landrace collection (Table 3). The late flowering landraces were also well represented in the landrace collection (18.6%). Western Africa is more than twice represented in this category of material than in the landrace collection. About 65% of late landraces come from this region. At the opposite, few late landraces are originated from the Indian subcontinent, Central or Southern Africa and Northern America. No late material was acquired from Eastern Asia and Oceanic.

#### Plant height

Plant height ranks from 80 to 655 cm. It presents a modal distribution slightly skewed towards the lower values. Traditionally are considered short the plants for which the height is smaller than 150 cm and tall when the height is greater than 450 cm. In our study, the short plant landraces are rare in the landrace collection (2.1%). Among this class, Northern America is near six times more represented than in the landrace collection. Eastern Asia and Oceanic and Southern Africa also are over-represented in this category of



landraces (about four times more than in the landrace collection). This later has the highest contribution to the shortest landraces. The tallest plants account for nine percent of the landrace collection. Western Africa mainly contributes to this type of landraces (factor 1.7 that means 46.3% of tall plants from this region). At the opposite, tall Southern African material is very rare.

#### Peduncle exertion

The size of the peduncle exertion varies from null to 71 cm. The absence of exertion concerns 6.1% of the landraces within the landrace collection. The representation of the Indian subcontinent within such material is significantly increased (factor 1.6) in comparison to its representation within the landrace collection and this region has a great contribution (27.2%). Excluding the null modality, peduncle exertion has a modal distribution. Among the landraces with peduncle exertion longer than 30 cm (6.5% of the landrace collection), the frequency of the material from Eastern Africa is increased by factor 1.6 (54.8% of such landraces are from this region). At the opposite, Indian subcontinent, Central Africa, Southern Africa and Eastern Asia and Oceanic have less than half of their representation in the well exerted material than in the landrace collection, and all together contribute for less then 16.7% of the landraces with long peduncle.

#### Seed weight

The 100 seed weight presents a modal distribution slightly skewed towards the higher values. The light seeded landraces (100 seed weight lower than 2 grams) correspond to 12.3% of the landrace collection. For this class, the representation of Central Africa and Northern America is more twice than those found in the landrace collection while Western and Southern Africa are under-represented. The heaviest seed landraces (100 seed weight higher than 4 grams) represent 16.8% of the landrace collection. The part of Western Africa is increased by 1.8 relative to its part in the landrace collection. On the contrary, a very small proportion of landraces from Central (1.1%), Southern Africa (2.2%) and from Eastern Asia and Oceanic (0.9%) present such criterion.

#### Leaves pigmentation

Leaves are mainly pigmented (97.6%). As a result, no specific change in the geographical distribution is found. However, differences are found for the tan plants. The proportion of the Northern America and Central Africa is increased by a factor 2.6 and 1.9 respectively. At the opposite, such material is less represented in Southern Africa (factor 0.1) and in Western Africa, Eastern Asia and Oceanic (factor 0.5).

#### Midrib pigmentation

Midrib pigmentation is recorded with four modalities: white, dull, yellow and brown color. No particular geographic tendency is drawn for the commonest color i.e. white that represents 82.4% of the landrace collection. The dull midrib accounts for 16.8% of the landrace collection. Southern Africa as Northern America is more than two folds more represented than they do in the landrace collection. Western Africa is under-represented in this class of material (factor 0.4). Yellow midrib is represented in the landrace collection at a frequency lower than 1% and more than the half part (59.2%) comes from Eastern Africa. The brown midrib is very rare (0.1% of the landrace collection). It is represented by accessions acquired from Northern America (factor 22.5), Central Africa (factor 3.4) and Eastern Africa. The two first regions each contribute to 38.5% to this type of material.

#### Glumes color

Two loci control the glumes color (Doggett, 1988). The non-pigmented glumes are either sienna or mahogany (30.6% and 11.1% of the landrace collection, respectively). Pigmented glumes are either black purple or red purple (42.3% and 16% of the landrace collection, respectively). Among the sienna glume landraces, the representation of Indian subcontinent and Southern Africa is increased by a factor 2.0 and 1.7 respectively. Northern America is near two folds more represented in the class of mahogany glumes than in the landrace collection. For black purple glume landraces, only a reduced contribution of the Indian subcontinent (factor 0.4) is noted. Eastern Asia and Oceanic is 1.7 fold more represented within the pigmented red purple glumes than in the landrace collection.

#### Seed covering

The seed covering by the glumes is recorded at maturity and is encoded with three classes from uncovered seed to completely covered seed. Only 5% of the landrace collection have uncovered seeds. An increased representation of Western Africa and Central Africa is found in this category of material. Conversely, landraces with uncovered seeds are under-represented in Southern Africa, Eastern Asia and Oceanic, and Northern America while they are very rare in Eastern Africa. Fully covered seeds account for 2.8% of the landrace collection. The representation of Northern America is 8 folds those of this region in the landrace collection. Nevertheless, this region contributes to only 14% of the covered seed landraces.

#### Panicle shape

Panicle shape and compactness are described from a compact both oval and elliptic panicle, to a very loose with either dropping or stiff branches. Landraces with loose panicle and stiff branches are largely represented in the landrace collection (47%). The representation of the Eastern Asia and Oceanic among this type of landraces is increased by 50% in comparison to the landrace collection. Conversely, the representation of the Indian subcontinent is decreased by 50%. Compact panicle landraces are as well commonly found in the landrace collection (45%). The Indian subcontinent is more frequent in this category than in the landrace collection (factor 1.5). At the opposite, Central Africa and Eastern Asia and Oceanic are under represented (factor 0.5 and 0.4). Loose panicles with dropping branches are less common in the landrace collection (8%). Among this category, the Central Africa is 1.8 fold more represented than in the landrace collection. The opposite situation is observed for Southern Africa.

#### Endosperm texture

Endosperm texture is coded from a completely corneous endosperm to a completely starchy endosperm. Only 4% of the landrace collection present a completely corneous endosperm. Western Africa contributes mainly to this type of landraces (factor 2.1) and the representation of Central Africa also is increased (factor 1.7). Conversely, Eastern Africa is poor in this type of landraces. The landraces with completely starchy endosperm represent 13.5% in the landrace collection. The Eastern Africa is more represented in this category than in the landrace collection (factor 1.7). At the opposite, the Indian subcontinent, Southern Africa, Northern America and Eastern Asia and Oceanic contribute to a small part to this type of material (7% over these four regions).

#### Testa

The testa is noted presence or absence. Around 1/3 of the landrace collection is characterized by the presence of testa. Eastern Asia and Oceanic is more than two folds



represented within this category of landraces than in the landrace collection. The opposite case is observed for Indian subcontinent. Absence of testa is noted in all regions.

#### Grain color

Grain color results from several interacting factors such as pericarp color, pericarp thickness, presence of testa, color and thickness of the testa (Rooney, 1980 in (Rooney and Miller, 1982)). The most common grain color in the landrace collection is brown (28.9%). Eastern Asia and Oceanic is 2.8 folds more represented among brown grain landraces than it does in the landrace collection. Conversely, the Indian subcontinent is less represented. The white grains and the straw grains are both well represented in the landrace collection (20%). Indian subcontinent has a higher representation among white grains while the opposite is noted for the Eastern Asia and Oceanic and Eastern Africa. Indian subcontinent well contributes to the straw grains. Red grains represent 15.9% of the landrace collection. Southern Africa is twice more represented among the red seeds landraces than over the landrace collection. Chalky white and gray grains account for 10% of the landrace collection. The main origin of this material is Eastern Africa (64% of these landraces). At the opposite, Indian subcontinent, Southern Africa and Eastern Asia and Oceanic have a representation reduced by three to five folds than it does in the landrace collection. Yellow grains contribute to a small part in the landrace collection (5.4%) and they were mainly found in the Indian subcontinent and Western Africa. They are very rare in Central Africa, Southern Africa and Eastern Asia and Oceanic. The landraces whose panicle is compounded by a mixture of red and white grains are the less represented (0.1%) in the landrace collection. These landraces were acquired only from Western Africa (75%) and Eastern Africa.

Racial tendencies are assessed to describe the morpho-agronomical variability in relation to the taxonomic entities defined through the racial classification. For this study, only the races present at more than 1% in the landrace collection were considered. The distribution of particular phenotypes for each of the nine most common races is compared to the distribution of each race in the landrace collection (Table 4). The highest and significant differences are then highlighted.

#### Days to flowering

The kafir race is more four folds represented among early landraces than in the landrace collection while it does not includes late material. An opposite situation is noted for the guinea race. Indeed, this race less contributes to the early landraces than it does to the late ones (3.5% and 45.1%, for the early and the late material, respectively). However, four races (DC, durra, caudatum and GC) give more than 60% of the early landraces while guinea and caudatum contribute both for 61.8% of the late landraces.

#### Plant height

The frequency of kafir landraces within the short plants is four times higher than its frequency in the landrace collection. However, kafir race contributes to only 9.2% of short plants. The proportion of CB and GC races is also increased among short plants (factor 2.8 and 2.3, respectively) and both together, they include about 34% of short plants. Conversely, the representation of guinea race is highly reduced and those of durra and DB races too but in a less proportion. Among the tall plants, the guinea race is three times more represented than in the landrace collection and it gives 53.4% of such plants. Tall plants were not found in the kafir race.



#### Peduncle exertion

Absence of peduncle exertion is mainly found within the durra race. More than 65% of these landraces belong to durra, caudatum or GC. Conversely, the guinea race is two fold less encountered among this type of landraces than in the landrace collection. The representation of DC and bicolor is higher among the long peduncle landraces than in the landrace collection.

#### Seed weight

The kafir race is very poorly represented both within light and heavy seed landraces. The representation of the bicolor, guinea and DB races is higher (4, 1.6 and 1.6 folds) among the light seeded landraces than it does over the landrace collection. The frequency of DC and durra races is significantly reduced in this category of landraces. DC race is twice more frequent among heavy seeded landraces than in the landrace collection. At the opposite, bicolor has a small contribution to the heavy seeded landraces.

#### Leaves pigmentation

Within the pigmented leaf landraces, no major change in the racial distribution is found. However, among the tan plants, the distribution of the GC race is three times increased relative to its presence in the landrace collection, while the representation of the durra, kafir and DB races is significantly reduced (factor 0.2, 0.3 and 0.2, respectively).

#### Midrib color

The racial distribution among the white midrib landraces is quite similar to those over the landrace collection. The presence of kafir, CB and GC races is significantly higher among dull midrib landraces than in the landrace collection. At the opposite, few dull midrib landraces belong to the guinea race (3.5%). The representation of bicolor or durra race is more important within yellow midrib landraces than in the landrace collection. These two races contribute to 54% of the yellow midrib plants. Such plants are not found in the kafir race. The CB and GC races are more represented among the rare brown midrib landraces (factor 4.5 and 2.9, respectively) while bicolor, kafir and DB races do not include such material.

#### Glume color

The frequency of durra, kafir and DB races is increased within the sienna glume landraces relative to their own frequencies in the landrace collection but their contribution to this phenotype is unequal (36.8%, 4.3%, and 9.1%). For mahogany glume, the representation of DB race is higher than in the landrace collection while those of caudatum and kafir races is reduced (factor 0.5). However, 11.9% of the corresponding landraces belong to the caudatum race versus 9.6% to the DB race. Among the red purple glume landraces, the frequency of caudatum and GC races is more important than in the landrace collection and those of durra and DB races are decreased. The racial distribution is not greatly disturbed within the black purple glume landraces in comparison to the racial distribution in the landrace collection.

#### Seed covering

The uncovered seed landraces are almost all found in the guinea race (99.1%). At the opposite, the bicolor race is 21 times more represented among the fully covered seed landraces than in the landrace collection. Only very rare such landraces with fully covered seed belong to the caudatum, guinea or DC races while they are completely absent within the kafir and GC races.

### Panicle shape

Generally, all races but durra are well represented among the landraces with loose panicle and stiff branches. The same remark is also made for the loose panicle with droopy branches. For this later case bicolor and guinea frequencies are highly increased in comparison to those found in the landrace collection (factor 2.4 and 3, respectively) while the representation of both durra and kafir are strongly reduced (factor 0.1 for both). Within the compact panicle landraces, the representation of the durra race is almost double than it does in the landrace collection. On the contrary, very few compact panicle landraces belong to the bicolor and guinea races.

### Endosperm texture

Corneous endosperm landraces belong mainly to the guinea race (84.9%). Among the starchy endosperm landraces, the representation of caudatum is more than twice higher those of this race in the landrace collection. At the opposite, the representation of bicolor, durra, guinea and kafir races are decreased by more than 50% in comparison to their distribution in the landrace collection.

### Testa

Among the landraces with a testa, the representation of the bicolor and caudatum race is increased in comparison to their distribution over the landrace collection by a factor 1.5 and 2, respectively. Conversely, this type of landraces is under-represented within the durra, guinea and kafir races.

### Grain color

The bicolor, caudatum and CB races are more represented within the brown grain landraces than in the landrace collection, but their contribution is unequal (8.4%, 46.1% and 6.5%). On the contrary, the durra, guinea and kafir races are poorly represented in such type of material. Among the white grain material, the representation of the guinea race (41.1%) is increased by more than two folds in comparison to its distribution in the landrace collection. In the category of straw grain landraces, the frequency of durra race (43.7%) is about two times increased relative to its proportion in the landrace collection while those of bicolor, caudatum and CB races is reduced. Among the red grain landraces, although the representation of the kafir race is three folds higher than in the landrace collection, it is only of 7.2% among red grains. The representation of caudatum and DC races among the chalky white and gray grain landraces (43.5% and 28.9% respectively) is increased by factor 2.0 and 2.4 compared to their representation in the landrace collection. But, the frequency of the bicolor, durra, kafir, DB and guinea races is sharply reduced in this type of landraces. Among the yellow grains landraces, the representation of the durra (56.3%) and DC (26.4%) races is more than twice increased relative to their proportion in the landrace collection while those of bicolor, caudatum, guinea, kafir and GC races is highly reduced. The landraces with both white and red grains were rarely found in the landrace collection and are absent in five races (bicolor, kafir, CB, DB and GC). The DC and caudatum races are in a higher proportion than in the landrace collection and contribute both to 62.5% of these landraces.

### Genetic diversity

Considering only the races presented at more than 1% of the landrace collection, and the four regions of Africa plus the Indian subcontinent, Shannon-Weaver diversity indices were calculated per race, region, as well as per race within regions (Table 5). They



rank from 0.54 to 0.74 for Southern Africa and Central Africa, respectively. The less diverse race is kafir with SDI = 0.46. The most diverse races are GC and CB (similar SDI of 0.72 and 0.70).

Shannon indices calculated for race within regions show the lowest value for the kafir race in the Indian subcontinent (SDI = 0.15). This low index could be explained by the size of the sample (three accessions). However, the highest index for kafir race (SDI = 0.51) is obtained for three accessions but coming from Western Africa. Thus, if the Indian sample can be assumed as a biased representation, and perhaps does not well reflect the diversity of the kafir race in India, the one of Western Africa is relatively good, as the diversity of these accessions is similar to the global diversity of the race. The highest diversity index is found in the GC race from Eastern Africa and Indian subcontinent (SDI = 0.74 and 0.72). These similar values were obtained despite a large difference in sample size (808 accessions from Eastern Africa versus 95 from India).

## DISCUSSION

Our study shows an unequal geographical distribution of the entire landrace collection (landrace collection) between the regions. However, it is in good concordance with the distribution of the largest sorghum growing areas over the world. For example, Africa and Indian subcontinent represent 79.4% of the world's sorghum area (FAO, 1999) and 95.6% of the landrace collection. The origin of sorghum is supposed to have taken place in the Northeastern quarter of Africa some 8,000 years ago (Wendorf et al., 1992). Therefore, it is not surprising that Eastern Africa was the most represented region (34.3% of the landrace collection). Another factor that is implied in the actual distribution of sorghum growing areas is its own history. Indeed, the spreading of sorghum through Africa as well as migratory events out of Africa led to several secondary centers of diversification (Harlan & Stemler 1976, Ollitrault et al. 1989) among which South Africa that is poorly represented. Out of Africa, in a first step sorghum reach India 3,000 years BP then China following the silk trade routes by 2,000 years BP. Consequently, as expected, Indian landraces are largely represented in the landrace collection (17.3%) but unfortunately, Chinese ones only correspond to 1.2%. This representation of landraces from China is largely biased since Chinese collection started in the 50<sup>ies</sup> has reached 10,386 accessions, for which a large morpho-agronomic diversity and a large range of valuable resources for genetic improvement were found (Li & Li, 1997). However, at the molecular level, landraces were found less diverse than improved varieties produces with exotic germplasm, thus suggesting a relatively narrow sorghum genetic base in China (Yang et al., 1994).

The global phenotypic diversity as measured by Shannon-Weaver indices for races as well as for regions or races within regions supports the hypothesis of Harlan & Stemler (1976), and corroborates the pattern of diversity as observed by Ollitrault et al. 1989, Aldrich & Doebley (1992), Aldrich et al. (1992) and Deu et al. (1994, 1995). Furthermore, the Shannon-Weaver indices calculated here could be good indications for future sampling as the comparison between the diversity of a race within a region and the race or the region in the landrace collection can lead to keep in mind the sample size effect. As an example, for guinea race, Dégremont (1992) and Deu et al. (1994) observed diversity within-guinea partly due to specific Western African forms (namely guinea margaritifera). Thus, it was expected a highest global diversity in Western Africa. But such was not the case. Also, because of the sample size, differences between diversity indices for guinea from Eastern



Africa (1.4% of the guinea in the landrace collection) and the guinea from West Africa (61.6%), was not significant. So, despite the 2,405 guinea accessions from Western Africa, one can wonder whether the landrace collection well reflect the variability of Western African guinea.

Using the ICRISAT database, our study shows a wide range of phenotypic diversity (SDI = 0.74) of the landrace collection. When only regions corresponding to more than 5% of the landrace collection and that have been identified as sorghum diversification centers (Harlan & Stemler, 1976) are considered, several observations can be made.

In Eastern Africa, all phenotypes are found. However, uncovered grains and corneous ones are rare while the half part of long peduncle exertion, yellow midrib, starchy and/or chalky grain landraces is from this region. Our study reveals that 22% of the accessions from Eastern Africa are early and only 1% is short. Gebrekidan (1982) reported that early and short plants are not widely found within Ethiopian sorghum collections. In their study, Ayana and Bekele (1999) concluded that accessions from Eritrea could be a good source of early maturing and short plant height genes. As it is underlined by Ayana and Bekele, it is possible that Eritrea was not well represented in the landrace collection. Therefore, despite the good representation of a region in the landrace collection, it is not sure that countries or areas within this region are correctly sampled.

All phenotypes, but the mixture of white and red grain color, are also found in Central Africa. This region contributes for more than 20% to the tan plants, brown midrib, loose panicle with drooping branches and light seed landraces. However, it poorly represents the heavy seeded and/or the yellow seeded landraces with less than two percent of those characteristics. Consequently to this high variability, the phenotypic diversity of this region is the most important and as high as in the landrace collection.

In Southern Africa, the phenotypic diversity is the most reduced (SDI = 0.54). Brown midrib and the mixture of white and red grain color landraces are not found in this region and late flowering, tall, tan, and yellow grain plants are rare (less than 10% of southern African accessions). On the contrary, more than 20% of the early flowering and short landraces are from this region.

Only brown midrib landraces are not found in Western Africa. This region contributes highly to the late landrace (65.8%), tall plants (about 50%), heavy seeded, uncovered, corneous and/or the mixture of white and red grain landraces. Interestingly, while guinea is the prevalent race in this region (40%) and although this race is represented with 20% light seeded material and only 8% heavy seeded material, 6% and 30% of the landraces found in Western Africa are light seeded and heavy seeded, respectively.

The Shannon-Weaver diversity index indicates a good global phenotypic diversity (SDI = 0.65) in India although it is lesser than those found in all the regions of Africa but Southern region. For Appa Rao et al. (1996) the diversity observed is important and higher among different states of India than within a state. Neither brown midrib nor red plus white grains landraces were collected in this region. However, the Indian subcontinent contributes to more than 1/5 of the landraces characterized by any of the following traits: null peduncle exertion, dull midrib, sienna glumes, compact panicle, absence of testa and light grains (i.e. white, yellow or straw grains).

The racial representation of the landrace collection is unbalanced too; out of the 15 sorghum races, five represent each one between 10.7% and 21.8% of the landrace collection. Also, although highly diverse, bicolor race is poorly represented in the landrace collection (only 3.2% of the landrace collection). As main results, the specific geographical distribution of races is in agreement with the domestication pathways (Harlan & Stemler, 1976; Ollitrault et al., 1989). Six of the nine more frequent races (those nine with frequency higher than 1%) were mainly collected from Eastern Africa. Furthermore, some of the basic races have their main distribution in their region of diversification. For example, caudatum is mainly found in Eastern Africa, guinea in Western Africa and kafir in Southern Africa. For durra race, this concordance between distribution and diversification center will depend upon the domestication hypothesis; either in the Indian subcontinent (Harlan & Stemler, 1976) or in Eastern-Africa (Ollitrault et al., 1989).

From our analyses, the main specific characteristics of each basic race are obvious. For example, within fully covered grains, it is highly probable to identify a bicolor (59% of bicolor landraces present such a peculiarity). In the same manner, 50% of bicolor landraces have light seeds. At least these two traits make this race considered as primitive and not really cultivated (in Africa or India) although it is present in all regions. As a result of this characteristic, bicolor race has several uses: fodder crop, syrup or molasses from its sweet stems and manufacture of beer from the grains (Harlan & de Wet, 1972).

Caudatum race is one of the most cultivated races. Grains with testa are well represented by the caudatum race (44%) and they also are the main characteristic of the caudatum race (69%). This particular character required fermentation processes prior to use it for culinary purpose. Also, its mainly floury endosperm (35% of the caudatum race) gives to caudatum best starch yield. As described by Harlan and de Wet (1972), the range of panicle compactness of this race is important. It makes caudatum sorghum well adapted to the driest zones as well as high rainfall areas. On the same manner, early and late flowering as well as short and tall plants are found within caudatum.

At the opposite, some races appear highly specialized. For example, by its compact panicle (86% of the durra race), durra sorghum is largely adapted to the driest regions. Also this race contributes for 29.3% to heavy seeds, and yellow (56.3%) or straw grain (43.7%) with no testa (29.4%) as required for good quality flour.

Guinea sorghums are known to be the oldest cultivated sorghum and mainly adapted to wettest zones (27.8% of the small-seeded types, 99.1% of the uncovered seeds and 51.5% of the loose droopy panicle) (Harlan & de Wet, 1972). In agreement with Irvine (1969, in Stemler et al., 1975), the guinea race mainly yields superior and white flour (47% of the guinea race). An uncovered seed is a major characteristic of the guinea race (28% of the guinea) and 99.1% of landraces with uncovered seeds belong to the guinea race. This character gives a free threshing quality. The corneous endosperm is also an important characteristic of guinea race (19% of the guinea race) while this character is rather rare in the landrace collection (4%). Guinea sorghum contributes to this phenotype for 84.9%. This trait confers to the guinea race a higher yield during dry milling.

Only kafir race, mainly cultivated under the temperate latitudes and characterized by early flowering plants, have a reduced morpho-agronomical variability ( $SDI = 0.46$ ). For five characters, some extreme phenotypes do not exist. For example, none kafir plant is late flowering, nor tall, nor with fully covered or uncovered grain. Kafir race is mainly early



flowering (94% of the kafir landraces) with dull midrib (37%) that gives sweet stalks mainly used for chewing.

All these morpho-agronomic descriptions are in agreement with the Harlan and de Wet race classification. However, rare characteristics within some races could be identified. Although corneous grain are mainly found in the guinea race it is interesting for the breeding consideration to know that few caudatum and durra landraces have corneous grains (0.06 and 0.3%, respectively), as well as rare guinea landraces have starchy grains (1%). On the same manner, early and late flowering as well as short and tall plants are found within guinea races despite the fact that early and short plants remain rare within this race (4% and 0.3%, respectively). Interestingly, despite their main tendency, guinea and durra races also present a wide range of variation in the panicle compactness, as shown with only 2% of guinea landraces with a compact panicle and 0.6% of durra with a loose and drooping branches one.

This study also reveals that on average, a good recombination occurs in hybrid forms. For example, 0.06% of caudatum and 19% of guinea have corneous grains versus 0.8% of GC. Also 35% of caudatum landraces and 1% of guinea ones have starchy grains versus 13% of GC forms. Globally, for endosperm texture the distribution for guinea is 20% in the two extremes classes, for caudatum it is 35% and for GC it is only 14%. The remaining 86% of GC is in the intermediate class. For breeding purpose, the absence of genetic obstacles was found for recombination of genetic components of both productivity and grain quality in guinea x caudatum crosses (Rami et al., 1998). The flowering pattern in kharif season is bimodal for caudatum (mode values equal to 90 and 130 days) and for GC (70 and 130 days) but unimodal for guinea (130 days). Such good capability in recombination between these two races explains the relative abundance of the GC hybrid forms in the landrace collection.

Also, from this study we found poorly represented kafir hybrids forms. This could be the result of the relatively recent domestication of the kafir race and its limited diffusion. The few GD intermediates could also be surprising. Indeed their two basic races (guinea and durra) are sympatric in some regions of Western Africa. One possible explanation of this rare representation could be due to their own morphological characteristics relevant on adaptations for antagonist climatic conditions. Thus it leads to difficult recombination to obtain and select "coherent" hybrid forms with good and interesting characteristics.

In conclusion, our study highlights the great potentiality of the cultivated sorghum. The agro-morphological diversity is important both within regions and within races. However, due to the sample size we analyzed (22,473 landraces) and consequently to the wide main regions considered (ten over the world), our results give only the main tendencies and pointed out some regions and/or countries that could be not well enough represented in the landrace collection. Interestingly, the frequent intermediate forms GC, CB and DC, and the two basic races bicolor and caudatum present the highest levels of phenotypic diversity. This result supports the fact that many breeding programs are based upon inter-race hybrid combinations. However, due to the level of phenotypic diversity we found, a better knowledge of the germplasm maintained in the landrace collection would be of great value for breeding programs. Our work also indicates that national collections can play an important role in the knowledge and the preservation of the germplasm. Such



collections have to be evaluated as it was done for India (Appa Rao et al., 1996) and Ethiopia and Eritrea (Ayana & Bekele, 1999). In conclusion, while many efforts are necessary to collect and to preserve for international long term gene banks, the preservation in national gene banks associated to a free access of data bases should be developed. Studies of both the global collection and specific national ones would be of great interest since the data could be efficiently used for breeding programs as well as for developing efficient management strategies for the germplasm collections.

## REFERENCES

- Appa Rao, S., K.E. Prasada Rao, M.H. Mengesha & V. Gopal Reddy, 1996. Morphological diversity in sorghum germplasm from India. *Gen Res and Crop Evol* 43: 559-567.
- Ayana, A. & E. Bekele, 1999. Multivariate analysis of morphological variation in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from Ethiopia and Eritrea. *Gen Res and Crop Evol* 46: 273-284.
- Dahlberg, J.A. & M.S. Spinks, 1995. Current status of the US sorghum germplasm collection. *ISMN* 36: 4-12.
- Doggett, H., 1988. *Sorghum*. 2<sup>nd</sup> edition Longman Burnt Mill, Harlow, Essex, England.
- FAO, 1999. <http://apps.fao.org/>
- Gebrekidan, B., 1982. Utilization of germplasm in sorghum improvement. In: ICRISAT (Ed), *Sorghum in the eighties: proceedings of the international symposium on sorghum*, pp. 335-345. ICRISAT, Patancheru, A.P., India.
- Harlan, J.R. & J.M.J. de Wet, 1972. A simplified classification of cultivated sorghum. *Crop Sci* 12: 172-176.
- Harlan, J.R. & A. Stemler, 1976. The races of sorghum in Africa. In: J.R. Harlan, J.M.J. de Wet, A.B.L. Stemler (Eds), *Origins of African Plant Domestication*, pp. 465-478. Mouton publishers, The Hague-Paris.
- House, L.R., 1985. A guide to sorghum breeding. ICRISAT, Patancheru, A.P., India.
- Hutcheson, K., 1970. A test for comparing diversities based on the Shannon formula. *J Theor Biol* 29: 151-154.
- Li, Y., S. Wu, Y.S. Cao & X. Zhang, 1996. A phenotypic diversity analysis of foxtail millet (*Setaria italica* (L.) P.Beauv.) landraces of Chinese origin. *Gen Res and Crop Evol* 43: 377-384.
- Li, Y. & C. Li, 1997. Phenotypic diversity of sorghum landraces in China. *International Conference on Genetic Improvement of Sorghum and Pearl Millet*, pp. 659-668.
- Mengesha, M.H. & K.E. Prasada Rao, 1982. Current situation and future of sorghum germplasm. In: L.R. House, L.K. Mughogho, J.M. Peacock & J.V. Mertin (Eds), *Sorghum in the Eighties: Proceedings of the international symposium on sorghum*, pp. 323-334. ICRISAT, Patancheru, A.P., India.
- Ollitrault, P., M. Arnaud & J. Chantreau, 1989. Polymorphisme enzymatique des sorghos. II - Organisation genetique et evolutive des sorghos cultives. *L'agronomie tropicale* 44 (3): 211-221.
- Rami, J.-F., P. Dufour, G. Trouche, G. Fliedel, C. Mestres, F. Davrieux, P. Blanchard & P. Hamon, 1998. Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench). *Theor Appl Genet* 97: 605-616.
- Rooney, L.W. & F.R. Miller, 1982. Variation in the structure and kernel characteristics of sorghum. *Proceedings of the International symposium in sorghum grain quality*, pp. 143-162. ICRISAT, Patancheru, A.P., India.
- Stemler, A.B.L., J.R. Harlan & J.M.J. de Wet, 1975. Caudatum sorghums and speakers of chari-nile languages in Africa. *Journal of African History* XVI (2): 161-183.
- Stenhouse, J.W., K.E. Prasada Rao, V. Gopal Reddy & S. Appa Rao, 1997. Sorghum. In: D. Fucillo, L. Sears, & P. Stapleton (Eds), *Biodiversity in trust*, pp. 292-308. Cambridge University Press, Cambridge, UK.
- Wendorf, F., A.E. Close, R. Schild, K. Wasylkowa, R.A. Housley, J.R. Harlan & H. Krolik, 1992. Saharan exploitation of plants 8,000 years BP. *Nature* 359: 721-724.
- Yang, G.P., M.A. Saghai Maroof, C.G. Xu, Q. Zhang & R.M. Biyashev, 1994. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Mol Gen Genet* 245: 187-194.

Table 1: Distribution of the landrace collection maintained at ICRISAT into 10 regions. The number of accessions held and their percentage of the regional collection is given. Within a region, countries which accounted for less than 5% were grouped together. Regions and countries are listed in order of relative representation based upon the number of accessions (percentage in brackets as a sub-total).

Region	Country	Number of accessions	Percentage within a region
Eastern Africa	Yemen	1902	24.7
	Ethiopia	1831	23.8
	Sudan	1826	23.7
	Kenya	868	11.3
	Uganda	853	11.1
	Somalia	409	5.3
	Egypt	12	0.2
	sub-total:	7701	(34.3)
Western Africa	Cameroon	2297	37.8
	Nigeria	1123	18.5
	Mali	629	10.4
	Burkina Faso	537	8.8
	Niger	395	6.5
	Togo, Senegal, Benin, Chad, Ghana, Sierra Leone, Gambia, Cape Verde Islands, Côte d'Ivoire	1092	18.0
	sub-total:	6073	(27.0)
Indian Sub-continent	India	3893	98.5
	Sri Lanka, Pakistan, Bangladesh, Nepal	60	1.5
	sub-total:	3953	(17.6)
Central Africa	Zimbabwe	978	38.1
	Tanzania	499	19.5
	Malawi	388	15.1
	Zambia	329	12.8
	Rwanda	146	5.7
	Burundi, Mozambique, Zaire, Madagascar, Angola, Central African Republic, Congo	224	8.7
	sub-total:	2564	(11.4)
Southern Africa	South Africa	566	47.7
	Lesotho	260	21.9
	Swaziland	183	15.4
	Botswana	178	15.0
	sub-total:	1187	(5.3)
Eastern Asia and Oceanic	China	286	62.3
	Korea	76	16.6
	Australia	35	7.6
	Indonesia	28	6.1
	Japan, Myanmar, Philippines, Taiwan, Thailand, Papua New Guinea	34	7.4
	sub-total:	459	(2.0)
Northern America	USA	385	100.0
	sub-total:	385	(1.7)
Southern America	Mexico	25	39.7
	Argentina	17	27.0
	Guatemala	4	6.3
	Honduras, Jamaica, Cuba, El Salvador, Nicaragua, Venezuela, Dominican Republic, Haiti, Uruguay	17	27.0
	sub-total:	63	(0.3)

Table 1: Continued  
Arabic

Turkey	34	66.7
Iran	6	11.8
Afghanistan	4	7.8
Syria	4	7.8
Iraq, Saudi Arabia	3	5.9
sub-total:	51 (0.2)	
Northern Africa and Southern Europe		
Algeria	22	59.5
Morocco	7	18.9
Portugal	4	10.8
Spain	3	8.1
Cyprus	1	2.7
sub-total:	37 (0.2)	
total:	22,473	



Table 2: Racial and geographical distribution of accessions of the landrace collection conserved at ICRISAT (percentage of total collection in bracket).

	Regions										Total
	Eastern Africa	Western Africa	Indian Sub-continent	Central Africa	Southern Africa	Eastern Asia and Oceanic	Northern America	Southern America	Arabia	Northern Africa and Southern Europe	
Races											
bicolor	179	85	173	50	34	59	105	5	20	8	718 ( 3.2)
caudatum	2431	1495	84	546	131	158	38	7	2	1	4893 (21.8)
durra	1442	456	2442	76	55	6	35	21	8	4	4545 (20.2)
guinea	56	2405	560	801	44	4	33	2	1	1	3907 (17.4)
kafir	7	3	3	61	429	3	10	2	1	1	520 ( 2.3)
CB	224	175	48	118	75	69	37	6	4	4	760 ( 3.4)
DB	796	114	270	12	7	16	17	9	7	14	1262 ( 5.6)
GB	10	147	20	32	4	5	1	-	1	-	220 ( 1.0)
KB	7	1	8	3	9	25	1	-	-	1	55 ( 0.2)
DC	1658	609	191	104	45	68	28	5	3	2	2713 (12.1)
GC	808	561	95	670	151	40	68	4	1	1	2399 (10.7)
KC	36	2	2	28	127	1	5	2	1	-	204 ( 0.9)
GD	39	19	36	34	3	1	4	-	2	-	138 ( 0.6)
KD	7	1	21	23	55	4	1	-	-	-	112 ( 0.5)
GK	1	-	-	6	18	-	2	-	-	-	27 ( 0.1)
Total	7701	6073	3953	2564	1187	459	385	63	51	37	22473

Table 3: Distribution of the landrace collection into the seven most frequent regions

Trait	Modality	Total landrace collection	Eastern Africa		Western Africa		Indian Sub-continent		Central Africa	
		%	%	Ratio	%	Ratio	%	Ratio	%	Ratio
All traits			34.3		27.0		17.6		11.4	
Days to flowering	Early	21.7	34.4	1.0	3.4	0.1	14.9	0.8	11.6	1.0
	late	18.6	24.7	0.7	65.8	2.4	3.8	0.2	4.7	0.4
Plant height	short	2.1	18.8	0.5	8.2	0.3	10.3	0.6	18.2	1.6
	tall	8.7	21.7	0.6	46.3	1.7	16.1	0.9	15.2	1.3
Panicle exertion	null	6.1	30.5	0.9	18.0	0.7	27.2	1.6	15.5	1.4
	long	6.5	54.8	1.6	26.4	1.0	8.5	0.5	5.5	0.5
Seed weight	light	12.3	31.2	0.9	13.7	0.5	22.2	1.3	22.9	2.0
	heavy	16.8	39.9	1.2	48.2	1.8	9.4	0.5	0.7	0.1
Plant color	pigmented	97.6	34.1	1.0	27.4	1.0	17.6	1.0	11.1	1.0
	tan	2.4	42.2	1.2	13.1	0.5	16.6	0.9	21.9	1.9
Midrib color	white	82.4	34.8	1.0	30.5	1.1	16.0	0.9	10.7	0.9
	dull	16.8	30.7	0.9	10.5	0.4	26.2	1.5	15.0	1.3
	yellow	0.8	59.2	1.7	13.6	0.5	5.3	0.3	5.3	0.5
	brown	0.1	23.1	0.7	-	-	-	-	38.5	3.4
Glume color	sienna	30.6	28.5	0.8	14.5	0.5	34.5	2.0	10.9	1.0
	mahogany	11.1	33.5	1.0	26.5	1.0	18.0	1.0	12.0	1.1
	black purple	42.3	39.7	1.2	33.8	1.2	7.8	0.4	10.6	0.9
	red purple	16.0	31.6	0.9	33.5	1.2	10.7	0.6	14.1	1.2
Seed covering	uncovering	5.0	0.6	≅0	56.5	2.1	22.6	1.3	18.2	1.6
	covering	2.8	26.2	0.8	13.2	0.5	25.7	1.5	5.9	0.5
Panicle shape	compact	45.0	37.9	1.1	23.1	0.9	26.6	1.5	6.2	0.5
	loose stiff	47.0	32.6	1.0	29.7	1.1	9.2	0.5	14.9	1.3
	loose droop	8.0	23.7	0.7	33.4	1.2	16.2	0.9	20.0	1.8
Endosperm texture	corneous	4.0	2.2	0.1	56.6	2.1	16.0	0.9	19.7	1.7
	starchy	13.5	57.1	1.7	20.2	0.7	4.3	0.2	15.1	1.3
Testa	presence	34.0	47.7	1.4	24.3	0.9	5.2	0.3	11.0	1.0
	absence	66.0	27.4	0.8	28.4	1.1	24.0	1.4	11.6	1.0
Grain	brown	28.9	46.7	1.4	17.6	0.6	6.0	0.3	16.1	1.4
	chalky/gray	10.0	64.0	1.9	22.8	0.8	3.0	0.2	6.2	0.5
	red	15.9	35.5	1.0	36.5	1.4	7.1	0.4	6.8	0.6
	white	19.8	12.4	0.4	35.4	1.3	30.7	1.7	12.6	1.1
	yellow	5.4	35.0	1.0	36.2	1.3	25.5	1.4	1.8	0.2
	straw	20.0	22.0	0.6	24.3	0.9	34.8	2.0	12.4	1.1
	white&red	0.1	25.0	0.7	75.0	2.8	-	-	-	-

Table 3: Continued

Trait	Modality	Southern Africa		East Asia and Oceanic		North America	
		%	Ratio	%	Ratio	%	Ratio
All traits		1.7		2.0		5.3	
Days to flowering	early	3.7	2.1	8.2	4.0	21.8	4.1
	late	0.7	0.4	-	-	0.2	≅0
Plant height	short	10.0	5.9	9.4	4.6	23.4	4.4
	tall	0.3	0.2	0.2	0.1	0.1	≅0
Panicle exertion	null	1.3	0.8	1.7	0.8	4.9	0.9
	long	1.9	1.1	1.1	0.5	1.6	0.3
Seed weight	light	3.7	2.2	1.6	0.8	3.1	0.6
	heavy	0.7	0.4	0.1	0.1	0.7	0.1
Plant color	pigmented	1.6	1.0	2.1	1.0	5.4	1.0
	tan	4.4	2.6	1.1	0.5	0.4	0.1
Midrib color	white	1.3	0.8	2.0	1.0	4.2	0.8
	dull	3.4	2.0	2.0	1.0	10.9	2.1
	yellow	4.7	2.8	3.6	1.7	1.8	0.3
	brown	38.5	22.5	-	-	-	-
Glume color	sienna	0.9	0.6	0.9	0.4	8.9	1.7
	mahogany	3.3	1.9	2.8	1.4	3.0	0.6
	black purple	1.9	1.1	2.2	1.1	3.4	0.7
	red purple	1.5	0.9	3.4	1.7	4.8	0.9
Seed covering	uncovering	0.4	0.3	0.4	0.2	1.3	0.3
	covering	14.0	8.2	6.7	3.3	3.7	0.7
Panicle shape	compact	1.0	0.6	0.8	0.4	3.8	0.7
	loose stiff	2.4	1.4	3.1	1.5	7.4	1.4
	loose droop	2.1	1.2	2.5	1.2	1.7	0.3
Endosperm texture	corneous	1.2	0.7	2.2	1.1	1.8	0.3
	starchy	0.6	0.3	0.8	0.4	1.3	0.3
Testa	presence	2.1	1.2	4.7	2.3	4.2	0.8
	absence	1.5	0.9	0.7	0.3	5.8	1.1
Grain color	brown	2.6	1.5	5.7	2.8	4.2	0.8
	chalky/gray	1.4	0.8	0.7	0.3	1.6	0.3
	red	2.1	1.2	0.7	0.4	10.8	2.0
	white	1.1	0.7	0.6	0.3	6.3	1.2
	yellow	1.2	0.7	0.2	0.1	0.1	0.0
	straw	1.0	0.6	0.4	0.2	4.8	0.9
	white&red	-	-	-	-	-	-



Table 4: Racial distribution of the landrace collection for the various modalities of twelve morphological characters based upon the number of accessions that originated from a particular region as a percentage of all landrace accessions classified into a specific phenotypic modality. A index of representation was calculated as the ratio between the percentage of landrace accessions classified into a specific phenotypic class from a particular region and the representation of this region in the total landrace collection.

Trait	Modality	bicolor		caudatum		durra		guinea		kafir	
		%	Ratio	%	Ratio	%	Ratio	%	Ratio	%	Ratio
All traits		3.2		21.8		20.2		17.4		2.3	
Days to flowering	early	5.8	1.8	13.4	0.6	14.1	0.7	3.5	0.2	10.0	4.3
	late	1.3	0.4	16.7	0.8	7.3	0.4	45.1	2.6	-	-
Plant height	short	1.9	0.6	20.5	0.9	6.9	0.3	2.1	0.1	9.2	4.0
	tall	2.7	0.8	5.8	0.3	10.3	0.5	53.4	3.1	-	-
Panicle exertion	null	2.0	0.6	19.3	0.9	36.7	1.8	7.9	0.5	1.5	0.7
	long	4.9	1.5	15.6	0.7	12.7	0.6	22.9	1.3	0.8	0.4
Seed weight	light	12.9	4.1	23.5	1.1	8.5	0.4	27.8	1.6	0.4	0.2
	heavy	0.5	0.1	22.4	1.0	29.3	1.4	8.7	0.5	0.2	0.1
Plant color	pigmented	3.2	1.0	21.9	1.0	20.6	1.0	17.2	1.0	2.4	1.0
	tan	4.1	1.3	18.6	0.9	4.2	0.2	25.0	1.4	0.7	0.3
Midrib color	white	3.1	1.0	21.5	1.0	18.7	0.9	20.4	1.2	1.8	0.8
	dull	3.5	1.1	23.5	1.1	26.7	1.3	3.5	0.2	5.1	2.2
	yellow	10.1	3.1	14.8	0.7	43.8	2.2	1.8	0.1	-	-
	brown	-	-	15.4	0.7	7.7	0.4	7.7	0.4	-	-
	sienna	1.8	0.6	7.4	0.3	36.8	1.8	14.7	0.8	4.3	1.9
	mahogany	4.2	1.3	11.9	0.5	18.7	0.9	21.8	1.3	1.2	0.5
Glume color	black purple	4.1	1.3	30.1	1.4	12.4	0.6	17.6	1.0	1.3	0.6
	red purple	2.9	0.9	34.0	1.6	10.3	0.5	18.9	1.1	2.0	0.8
	uncovering	-	-	0.1	≅0	-	-	99.1	5.7	-	-
	covering	67.5	21.1	0.2	0.0	3.2	0.2	0.3	≅0	-	-
Panicle shape	compact	0.7	0.2	28.0	1.3	38.6	1.9	0.7	≅0	1.9	0.8
	loose stiff	4.8	1.5	16.9	0.8	5.8	0.3	27.6	1.6	3.0	1.3
	loose droop	7.5	2.4	15.2	0.7	1.4	0.1	51.5	3.0	0.3	0.1
	corneous	4.7	1.5	0.3	≅0	1.5	0.1	84.9	4.9	0.3	0.1
Endosperm texture	starchy	0.6	0.2	56.1	2.6	9.0	0.4	1.4	0.1	0.6	0.2
	presence	4.7	1.5	44.1	2.0	2.4	0.1	4.6	0.3	0.9	0.4
Testa	absence	2.4	0.8	10.3	0.5	29.4	1.5	24.0	1.4	3.0	1.3
	brown	8.4	2.6	46.1	2.1	3.3	0.2	3.2	0.2	0.5	0.2
Grain color	chalky/gray	0.1	≅0	43.5	2.0	1.5	0.1	5.3	0.3	0.4	0.2
	red	1.4	0.4	12.2	0.6	20.8	1.0	20.4	1.2	7.2	3.1
	white	1.3	0.4	6.1	0.3	20.3	1.0	41.1	2.4	2.8	1.2
	yellow	0.3	0.1	3.8	0.2	56.3	2.8	2.4	0.1	0.1	≅0
	straw	1.3	0.4	3.6	0.2	43.7	2.2	22.1	1.3	2.2	0.9
	white&red	-	-	37.5	1.7	25.0	1.2	12.5	0.7	-	-

Table 4: Continued

Trait	Modality	CB		DB		DC		GC	
		%	Ratio	%	Ratio	%	Ratio	%	Ratio
All traits		3.4		5.6		12.1		10.7	
Days to flowering	early	6.6	1.9	4.3	0.8	22.0	1.8	12.7	1.2
	late	2.9	0.9	8.0	1.4	8.9	0.7	5.1	0.5
Plant height	short	9.6	2.8	2.3	0.4	14.2	1.2	24.5	2.3
	tall	3.0	0.9	8.4	1.5	4.7	0.4	7.9	0.7
Panicle exertion	null	2.8	0.8	5.3	0.9	9.9	0.8	11.3	1.1
	long	4.4	1.3	5.3	0.9	23.3	1.9	8.4	0.8
Seed weight	light	4.7	1.4	8.8	1.6	3.9	0.3	7.2	0.7
	heavy	2.0	0.6	4.4	0.8	24.2	2.0	6.1	0.6
Plant color	pigmented	3.4	1.0	5.7	1.0	12.2	1.0	10.1	0.9
	tan	2.2	0.7	1.3	0.2	6.8	0.6	32.8	3.1
Midrib color	white	2.8	0.8	6.2	1.1	13.2	1.1	9.1	0.9
	dull	6.0	1.8	2.9	0.5	6.7	0.6	18.5	1.7
	yellow	4.1	1.2	6.5	1.2	9.5	0.8	8.9	0.8
	brown	15.4	4.5	≅0	≅0	15.4	1.3	30.8	2.9
Glume color	sienna	2.8	0.8	9.1	1.6	11.1	0.9	7.4	0.7
	mahogany	4.4	1.3	9.6	1.7	13.8	1.1	10.5	1.0
	black purple	3.7	1.1	3.5	0.6	14.0	1.2	11.0	1.0
	red purple	3.1	0.9	1.9	0.3	7.7	0.6	16.3	1.5
Seed covering	uncovering	-	-	-	-	-	-	0.8	0.1
	covering	4.4	1.3	22.1	3.9	0.2	≅0	-	-
Panicle shape	compact	1.9	0.6	3.8	0.7	14.8	1.2	7.2	0.7
	loose stiff	4.7	1.4	7.3	1.3	11.2	0.9	14.3	1.3
	loose droop	4.1	1.2	5.7	1.0	1.9	0.2	8.7	0.8
Endosperm texture	corneous	1.3	0.4	1.5	0.3	0.2	≅0	2.2	0.2
	starchy	1.8	0.5	4.1	0.7	16.2	1.3	9.9	0.9
Testa	presence	5.5	1.6	4.1	0.7	15.5	1.3	16.3	1.5
	absence	2.3	0.7	6.4	1.1	10.3	0.9	7.8	0.7
Grain color	brown	6.5	1.9	9.1	1.6	7.0	0.6	14.5	1.4
	chalky/gray	2.0	0.6	1.4	0.2	28.9	2.4	14.9	1.4
	red	3.4	1.0	4.7	0.8	16.4	1.4	8.8	0.8
	white	2.1	0.6	4.0	0.7	7.7	0.6	9.1	0.9
	yellow	2.1	0.6	4.5	0.8	26.4	2.2	1.8	0.2
	straw	1.2	0.4	5.3	0.9	8.0	0.7	8.5	0.8
	white&red	-	-	-	-	25.0	2.1	-	-

Table 5: Shannon-Weaver diversity index (standardized values  $\pm$  variance) for the nine most common races and five most represented regions. Effective number of each group is in brackets.

	Eastern Africa	Central Africa	Indian Sub-continent	South Africa	West Africa	Total
bicolor	0.64 $\pm$ 0.00 (179)	0.64 $\pm$ 0.01 (50)	0.64 $\pm$ 0.00 (173)	0.53 $\pm$ 0.01 (34)	0.71 $\pm$ 0.00 (85)	0.68 $\pm$ 0.00 (521)
caudatum	0.67 $\pm$ 0.00 (2431)	0.64 $\pm$ 0.00 (546)	0.60 $\pm$ 0.00 (84)	0.54 $\pm$ 0.00 (131)	0.59 $\pm$ 0.00 (1495)	0.67 $\pm$ 0.00 (4687)
durra	0.57 $\pm$ 0.00 (1442)	0.59 $\pm$ 0.00 (76)	0.51 $\pm$ 0.00 (2442)	0.49 $\pm$ 0.01 (55)	0.55 $\pm$ 0.00 (456)	0.56 $\pm$ 0.00 (4471)
guinea	0.70 $\pm$ 0.01 (56)	0.61 $\pm$ 0.00 (801)	0.55 $\pm$ 0.00 (560)	0.50 $\pm$ 0.01 (44)	0.58 $\pm$ 0.00 (2405)	0.62 $\pm$ 0.00 (3866)
kafir	0.36 $\pm$ 0.04 (7)	0.47 $\pm$ 0.01 (61)	0.15 $\pm$ 0.16 (3)	0.44 $\pm$ 0.00 (429)	0.51 $\pm$ 0.17 (3)	0.46 $\pm$ 0.00 (503)
CB	0.70 $\pm$ 0.00 (224)	0.64 $\pm$ 0.00 (118)	0.60 $\pm$ 0.01 (48)	0.52 $\pm$ 0.00 (75)	0.67 $\pm$ 0.00 (175)	0.70 $\pm$ 0.00 (640)
DB	0.62 $\pm$ 0.00 (796)	0.65 $\pm$ 0.03 (12)	0.63 $\pm$ 0.00 (270)	0.58 $\pm$ 0.05 (7)	0.59 $\pm$ 0.00 (114)	0.66 $\pm$ 0.00 (1199)
DC	0.64 $\pm$ 0.00 (1658)	0.67 $\pm$ 0.00 (104)	0.65 $\pm$ 0.00 (191)	0.51 $\pm$ 0.01 (45)	0.56 $\pm$ 0.00 (609)	0.67 $\pm$ 0.00 (2607)
GC	0.74 $\pm$ 0.00 (808)	0.67 $\pm$ 0.00 (670)	0.72 $\pm$ 0.00 (95)	0.48 $\pm$ 0.00 (151)	0.68 $\pm$ 0.00 (561)	0.72 $\pm$ 0.00 (2285)
Total	0.72 $\pm$ 0.00 (7601)	0.74 $\pm$ 0.00 (2438)	0.65 $\pm$ 0.00 (3866)	0.54 $\pm$ 0.00 (971)	0.70 $\pm$ 0.00 (5903)	0.74 $\pm$ 0.00 (20779)



Diversité morpho- agronomique des sorghos cultivés.

Grenier C, Deu M, Chantereau J, Bramel-Cox PJ, Hamon P

*Communication orale présentée au séminaire INRA (Méribel 1999)*

---

## DIVERSITÉ MORPHO-AGRONOMIQUE DES SORGHOS CULTIVÉS.

Cécile Grenier<sup>(1)</sup>, Monique Deu<sup>(1)</sup>, Jacques Chantereau<sup>(1)</sup>, Paula Bramel-Cox<sup>(2)</sup> et Perla Hamon<sup>(3)</sup>.

<sup>(1)</sup> CIRAD-CA-Calim, Avenue Agropolis - BP 5035 - 34032 Montpellier Cedex 1 ; <sup>(2)</sup> ICRISAT-GREP-GRD, Patancheru, 502 324 A.P. India ; <sup>(3)</sup> Université Montpellier III - route de Mende - 34199 Montpellier Cedex 5.

Le sorgho (*Sorghum bicolor* [L.] Moench) est une Andropogonée de la famille des Poacées qui représente un constituant majeur dans l'alimentation humaine des pays en voie de développement. Situé au 5<sup>ème</sup> rang mondial des productions céréalières, et au 2<sup>nd</sup> rang pour les pays d'Afrique, sa culture est particulièrement bien adaptée aux conditions tropicales semi-arides. Son aire de production d'environ 45 millions d'hectares, est répartie à 90% dans les pays en voie de développement; ces derniers contribuant pour 70% de la production mondiale, soit environ 44 millions de tonnes (ICRISAT, 1996). Les utilisations culinaires traditionnelles du sorgho varient considérablement entre les pays. Le sorgho peut avoir des utilisations très diverses: matériaux de construction, combustibles, ou encore la production de sirops, de sucre, d'alcool, et d'amidon. Dans les pays industrialisés, la culture du sorgho est surtout orientée vers l'alimentation du bétail.

Une classification raciale des sorghos cultivés, basée sur les caractères de l'épi et des épillets, a été proposée par Harlan et de Wet (Harlan et de Wet, 1972). Elle divise les sorghos cultivés en cinq races de base; bicolor, caudatum, durra, guinea et kafir, et 10 formes intermédiaires qui présentent des caractères morphologiques se rapportant à au moins deux races de base.

A l'heure actuelle, l'ensemble de la collection mondiale des sorghos cultivés (36 242 accessions) est conservé dans la banque de semences de l'ICRISAT en Inde. Le nombre sans cesse croissant des accessions collectées pose des problèmes d'ordre matériel aux responsables de telles collections. En effet, les gestionnaires de ces banques de germplasma doivent veiller à assurer la représentation et le maintien de la diversité, mais aussi se confronter aux difficultés de caractérisation et de régénération des entrées afin de satisfaire les demandes de semences (Brown, 1995; Marshall, 1989). Dans l'idée de faciliter la conservation des espèces cultivées et de leurs variétés sauvages apparentées, et afin d'optimiser l'utilisation des ressources génétiques, le concept de core collection a été développé en 1984 par Frankel. L'objectif en est de maintenir la diversité génétique, tout en réduisant le nombre des accessions devant être manipulées pour rester disponibles et accessibles aux utilisateurs (Brown, 1989).

Une étape préliminaire à l'application de stratégies d'échantillonnage consiste en l'étude de la diversité génétique de l'ensemble de la collection. Cette diversité peut notamment être estimée par des analyses de la variabilité morpho-agronomique, au travers d'études de la distribution des caractères morphométriques, et d'évaluation de la fréquence de modalités répertoriées à chacun des descripteurs discrets. Un travail descriptif visant à détecter l'existence d'une structure au sein

même de la collection, permettra par la suite d'établir une stratégie d'échantillonnage adaptée à un ensemble hiérarchisé. Dans ce travail nous nous proposons d'étudier la diversité génétique des sorghos cultivés en utilisant les caractères morpho-agronomiques, afin de mettre en évidence une éventuelle structuration au sein de la collection.

## MATÉRIEL

L'étude de la diversité morphologique de la collection est réalisée sur un ensemble d'accessions ne représentant que des variétés traditionnelles pour lesquelles les données morpho-agronomiques et passeport sont complètes. De plus, seules les accessions comprises entre 40 degrés de latitude Nord et 40 degrés de latitude Sud ont été retenues dans notre étude. En effet il est fortement probable qu'au delà de ces latitudes, les variétés cultivées ne soient pas des formes traditionnelles mais soient issues des programmes d'amélioration végétale. Cette réduction conduit à une taille de collection de base de 22 473 accessions. Les informations passeport disponibles sur cette collection portent sur la taxonomie botanique des accessions et leur origine géographique (Tableau 1). La diversité morpho-agronomique est étudiée à partir d'informations morphologiques collectées à l'ICRISAT depuis plus de 25 ans et portant sur 19 descripteurs dont dix variables continues et neuf variables discrètes (Tableau 2). Deux caractères morphométriques sont enregistrés pour les deux saisons typiques des zones tropicales semi-arides; la saison des pluies encore appelée la saison Kharif, de juin à décembre avec des jours de longue durée, et la saison sèche ou saison Rabi, de septembre à avril où les jours sont de courte durée.

## MÉTHODES

La diversité morphométrique des accessions de la collection de base est étudiée en utilisant les méthodes de statistiques univariées. Pour chaque variable continue sont obtenues les informations relatives aux intervalles de valeurs, moyenne et écart-type. La diversité des caractères discrets (coloration, forme, absence vs. présence de caractères, codification numérique) est estimée par une étude de distribution de fréquences pour chaque modalité de ces différentes variables.

La structuration raciale de la collection est étudiée par des analyses multivariées pour les variables continues, et des comparaisons des distributions de fréquences pour les variables discrètes (StatSoft, 1997). Dans ce but, un échantillonnage aléatoire de 500 accessions est réalisé au sein de chacune des cinq races de base. Une analyse de variance avec un test de Levene sur les caractères métriques vérifie l'homogénéité des variances. Les moyennes sont alors comparées par un test de Newman et Keuls. Pour les variables discrètes, le modèle log-linéaire teste la relation entre facteurs dans un tableau de fréquences d'ordre multiple. Ceci permet d'estimer la signification des interactions simples ou multiples avec les données taxonomiques et/ou géographiques (latitude ou continent). Un test de chi-2 donne la part observée qui peut être expliquée par le modèle.

## RÉSULTATS ET DISCUSSION

### *Diversité morpho-agronomique au sein de la collection de base*

La distribution géographique et latitudinale des variétés (Tableau 1 A) montre que plus des 2/3 de la collection sont d'origine africaine avec environ 15% en provenance du Cameroun. La seconde zone importante de distribution des sorghos se situe en Asie représentant ainsi 28% de la collection. Soixante-deux % des sorghos asiatiques, ont été collectés en Inde. Très peu de variétés traditionnelles sont originaires d'Amérique, d'Europe et d'Océanie, ceci relevant du fait



que les sorghos sont particulièrement adaptés et donc majoritairement cultivés au niveau des zones intertropicales semi-arides.

La contribution des cinq races de base et de leurs 10 formes intermédiaires à l'ensemble de la collection étudiée est inégale (Tableau 1 B). Chacune des races caudatum, durra et guinea représente 1/5 à un quart de la collection. Au total, près de 82% de la collection sont de race caudatum, durra, guinea, durra-caudatum et guinea-caudatum.

La distribution latitudinale n'est pas quelconque; elle est fonction des races. La race caudatum prédomine aux latitudes équatoriales (67%) alors qu'elle est faiblement représentée (3%) dans les zones les plus tempérées (latitude 4). Les races durra et guinea sont majoritairement cultivées aux latitudes tropicales (>56%) puis équatoriales, et ne sont que rarement présentes dans régions les plus tempérées (<1,3%). Enfin la race kafir et ses formes intermédiaires sont cultivées essentiellement au niveau des régions tempérées, avec seulement 2% recensés dans la zone équatoriale.

L'étude des distributions de fréquences des caractères continus permet d'apprécier la variabilité morphométrique des sorghos cultivés. Les variances notées pour les dates de floraison et hauteur de plante, sont relativement importantes et largement affectées par la variation saisonnière (Tableau 3). Ceci traduit une importante variabilité des sorghos pour les caractères végétatifs et les réactions à la longueur du jour. En effet, certaines variétés réagissent à l'augmentation de la photopériode en retardant leur floraison. Cette différence dans la photosensibilité résulte d'une interaction entre génotype et environnement. Les variétés identifiées comme précoces, pour lesquelles les dates de floraisons sont minimales pour les deux saisons, appartiennent majoritairement aux races caudatum et durra-caudatum. Les variétés tardives sont en majorité des accessions de race guinea, et guinea-caudatum. Il est intéressant de noter que cette catégorie de plantes ayant une floraison tardive est également bien représentée par les races caudatum et intermédiaires caudatum. Les variétés de petite taille sont essentiellement des formes caudatum (race caudatum et intermédiaires GC et DC); celles ayant un grand développement végétatif appartiennent en grande majorité à la race guinea. En règle générale, les analyses descriptives montrent que la collection mondiale renferme un bonne gamme de variabilité. Dans ce qui suit, seules les analyses portant sur quelques variables présentant un intérêt pour les programmes d'amélioration ont été détaillées (Fig. 1). Les dates de floraison enregistrées au cours de la saison des pluies sont caractérisées par une distribution bimodale avec d'une part des génotypes insensibles à la variation photopériodique principalement originaires des zones tempérées, et des variétés plutôt photosensibles essentiellement d'origine équatoriale. De même les variétés caractérisées par de grandes plantes à la saison des pluies sont essentiellement des accessions d'origine équatoriale et les petites plantes sont en majorité originaires des zones tempérées.

Les accessions présentant un tallage basale important ( $BT > 5$ ) sont essentiellement de race guinea et en majorité originaires de Sierra Leone (77%). Une telle information peut être utilisée pour orienter les prospections ou l'utilisation de variétés intéressantes pour les productions fourragères.

Les sorghos dont les panicules sont rondes et compactes sont des formes bien adaptées aux conditions sèches et sont surtout cultivés pour l'alimentation humaine. Ces accessions sont pour la plus grande part de race durra (77%) et caudatum (12%), et principalement d'origine indienne (33%) et somalienne (33%). Les variétés à panicule large et lâches facilitant le séchage, sont bien adaptées aux conditions humides. Elles appartiennent en majorité aux formes durra-bicolor (26%), guinea (24%), bicolor (15%), et caudatum (14%).

Les grains de petite taille ( $GRS < 1$  mm) semblent caractéristiques de la race bicolor (81%), et les gros grains ( $GRS > 4$  mm) sont essentiellement de formes durra et caudatum, respectivement 46%

et 15%. Pour la race guinea on constate une proportion non négligeable de gros grains (18% des accessions ont des grains de taille supérieure à 3,2 mm et environ 9% des accessions ont des grains tailles supérieures à 4 mm). Les grains de faible poids (SWT<1 gr) appartiennent essentiellement aux races guinea (55%) et bicolor (28%). La majorité des variétés avec des grains de poids très élevé (SWT>7 grs) appartient à la forme durra (98%) (race durra plus les intermédiaires DC, DB et DG).

Les sorghos dont les grains sont complètement vitreux représentent des variétés intéressantes de part leur moindre sensibilité aux attaques d'insectes et sont préférés pour la confection de porridges et des préparations culinaires de type riz (Rooney and Murty in ICRISAT, 1982). Ces variétés sont pour la majorité des cas de race guinea (85%). Les grains complètement farineux utilisés pour la fabrication des pains sont essentiellement des formes caudatum (82%).

Les variétés dont les grains pivotent à maturité sont faiblement recouverts par les glumes et ont de bonnes aptitudes au séchage dans des conditions climatiques humides. Ces variétés sont presque exclusivement de race guinea (99%). Le recouvrement complet du grain, caractère observé chez les sorghos sauvages, est essentiellement rencontré dans la race bicolor (67%) et ses formes intermédiaires (29%).

#### *Structuration de la collection de base*

L'étude descriptive de la diversité morpho-agronomique nous a permis d'appréhender l'existence d'une structuration raciale. Des analyses multivariées menées sur les caractères morphométriques et visant à discriminer la collection à partir des groupes raciaux ont montré une variance entre les races significativement supérieure à la variance intra-race (Tableau 4). A partir du test d'homogénéité de variance de Levène, neuf des dix caractères morphométriques ont montré des variances différentes, avec des valeurs de F statistiquement significatives ( $p < 0,01$ ). Un seul caractère fait exception, il s'agit de la hauteur de plante mesurée pendant la saison sèche ( $F = 2,26$  et  $p > 0,06$ ). Ainsi pour ce caractère, les moyennes des différentes races ont pu être comparées par le test de Newman et Keuls. Seules les races caudatum et kafir présentent des moyennes qui statistiquement ne sont pas différentes ( $p > 0,05$ ).

Globalement on peut donc dire que les cinq races de base sont bien différenciées par leur caractéristiques morphométriques. Il y a donc bien validation de l'importance du facteur racial dans la structuration de la collection en ne considérant que les cinq races majeures.

Les variables discrètes sont testées dans des modèles faisant intervenir des interactions entre les effets de race et d'origine géographique (continent et rang latitudinal). Tous les modèles proposés montrent que le phénotype global varie significativement en fonction de la race, associée ou non avec le continent d'origine ou la latitude. Une analyse de chi-2 a permis de déterminer la contribution des effets originaux et des effets des interactions au modèle retenu (Tableau 5). Ces tests montrent que dans pratiquement tous les cas, les écarts entre la distribution observée dans la collection et celle attendue dans le cas d'indépendance entre les races, les origines géographiques et les caractères phénotypiques, sont très significatifs. Les deux exceptions concernent l'interaction primaire entre le caractère de pigmentation foliaire et soit le continent ( $p > 0,21$ ), soit la latitude ( $p > 0,50$ ).

En résumé l'étude de la collection mondiale des sorghos cultivés montre l'existence d'une structuration raciale, et la possibilité d'une hiérarchisation éco-géographique liée à la photosensibilité de certains génotypes. Ces résultats nous ont permis de considérer une structuration à la fois raciale et éco-géographique (Grenier et al., soumis) de la collection de base. Depuis, plusieurs stratégies d'échantillonnage ont été appliquées comme cela a été suggéré par Spagnoletti Zeuli et Qualset, 1993 et Ortiz et al., 1998. Les travaux se poursuivent par la



constitution de core collections suivant différentes stratégies d'échantillonnage. La diversité génétique de ces core collections est évaluée en utilisant les caractères morpho-agronomiques, des marqueurs moléculaires, et les indices de richesse spécifique. Les analyses en cours devraient permettre de comparer l'efficacité des différentes stratégies d'échantillonnage sur la diversité génétique retenue dans les core collections. Des méthodes d'échantillonnage seront alors proposées tenant en compte les objectifs assignés à la core à constituer.

RÉFÉRENCES

Brown, A. H. D. (1989). The case for core collections. *In* The use of plant genetic resources, A. H. D. Brown, Frankel, O. H., Marshall, D. R., Williams, J. T., ed., Cambridge, pp.135-156.

Brown, A. H. D. (1995). The core collection at the crossroads. *In* Core collections of plant genetic resources., T. Hodgkin, Brown, A.H.D., van Hintum, T.J.L. , Morales, E.A.V, ed., John Wiley & Sons, Chichester, UK.

Grenier, C., Bramel-Cox, P.J., Hamon, P. (1999). Core collection of the genetic resources of sorghum: I. Stratification based on eco-geographical data. Submitted to *Crop Sciences*

Harlan, J. R., de Wet, J.M.J. (1972). A simplified classification of cultivated sorghum. *Crop Sci.*, 12, 172-176.

ICRISAT. (1982). Sorghum in the Eighties. *In* International symposium on sorghum. ICRISAT Center Patancheru, A.P., India.

ICRISAT. (1996). The world sorghum and millet economies - Facts, trends and outlook. ICRISAT & FAO.

Marshall, D. R. (1989). Limitations to the use of germplasm collections. *In* The use of plant genetic resources., A. D. H. Brown, Frankel, O.H., Marshall, D.R., Williams, J.T., ed., Cambridge, pp.105-122.

Ortiz, R., Ruiz-Tapia, E.N., Mujica-Sanchez, A. (1998). Sampling strategy for a core collection of Peruvian quinoa germplasm. *Theor Appl Genet*, 96, 475-483.

Spagnoletti Zeuli, P. L., Qualset, C. O. (1993). Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. *Theor Appl Genet*, 87, 295-304.

StatSoft (1997) . Statistica for Windows. USA, StatSoft.



Tableau 1: Information passeport sur l'ensemble de la collection mondiale des sorghos cultivés:  
 (A) pays d'origine, (B) distribution raciale et latitudinale.

(A)	Pays d'origine	Effectif	%			
	<b>Afrique</b>	<b>15652</b>	<b>69,7</b>			
	Cameroun	2297	10,2			
	Ethiopie	1831	8,2			
	Soudan	1826	8,1			
	Nigeria	1123	5,0			
	Autres	8575	38,2			
	<b>Asie</b>	<b>6329</b>	<b>28,2</b>			
	Inde	3893	17,3			
	Yémen	1902	8,5			
	Autres	2436	10,8			
	<b>Amériques</b>	<b>448</b>	<b>2,0</b>			
	USA	385	1,7			
	Autres	63	0,3			
	<b>Océanie</b>	<b>36</b>	<b>0,2</b>			
	Australie	35	0,2			
	Autres	1	0,0			
	<b>Europe</b>	<b>8</b>	<b>0,0</b>			

(B)	Races		lat.1	lat.2	lat.3	lat.4	Effectif
	Bicolor	B	206	260	69	183	718
	Caudatum	C	3284	1263	175	171	4893
	Durra	D	1334	2628	523	60	4545
	Guinea	G	1582	2177	95	53	3907
	Kafir	K	6	63	195	256	520
	Caudatum-Bicolor	CB	195	365	89	111	760
	Durra-Bicolor	DB	632	491	108	31	1262
	Guinée-Bicolor	GB	131	75	10	4	220
	Kafir-Bicolor	KB	4	15	28	8	55
	Durra-Caudatum	DC	671	1853	124	65	2713
	Kafir-Caudatum	KC	838	1200	255	106	2399
	Guinea-Caudatum	GC	9	53	89	53	204
	Durra-Kafir	DK	25	92	15	6	138
	Guinea-Durra	GD	1	43	50	18	112
	Guinea-Kafir	GK	1	5	18	3	27
			8919	10583	1843	1128	22473

avec lat.1 : (10°S-0°; 0°-10°N); lat.2 : (20°S-10°S; 10°N-20°N); lat.3 : (30°S-20°S; 20°N-30°N);  
 et lat.4 : (40°S-30°S; 30°N-40°N)

Tableau 2: Liste des descripteurs morpho-agronomiques

Variables continues:

• date de 50% de floraison des plants (jours depuis la date de semis)	FLR (Rabi) FLK (Kharif)
• nombre de talles racinaires	BT
• hauteurs de plantes (cm)	PHTR (Rabi) PHTK (Kharif)
• longueur du pédoncule (cm)	PEDEX
• longueur de la panicule (cm)	EHLG
• largeur de la panicule (cm)	EHWD
• taille des grains (mm)	GRS
• poids de 100 grains (grs)	SWT

Variables discrètes:

• coloration de la plante (2 couleurs)	PIG
• coloration de la nervure médiane (4 couleurs)	MCR
• compacité de la panicule (10 modalités)	EHCS
• coloration des grains (13 couleurs)	GRC
• coloration des glumes (13 couleurs)	GLC
• couverture du grain par les glumes (1: peu couvert - 5: couverture complète)	COV
• texture de l’albumen (1: grain vitreux - 5: grain farineux)	COR
• lustrosité des grains (2 modalités)	LUS
• couche brune (présence vs. absence)	SC

Tableau 3: Informations passeport (race et latitude) des variétés en fonction du cycle végétatif caractérisé à partir des dates de floraison (FLR et FLK) et des hauteurs de plante (PHTR et PHTK).

Cycle végétatif	Caractères	composition en races majoritaires	lat.1	lat.2	lat.3	lat.4
court	FLR≤50	DC 53% - C 17%	5%	84%	6%	3%
	FLK≤50	DC 22% - C 20% - B 20% - DB 11%	5%	52%	21%	22%
	PHTR≤100	C 41% - GC 14% - DC 13%	19%	45%	17%	18%
	PHTK≤100	DC 26%- C 21% - CB 10%- D 10%	21%	36%	18%	26%
long	FLR>120	G 51% - GC12%	35%	65%		
	FLK>160	C 32% - G 24% - GC 16% - DB 12%	60%	40%		
	PHTR>400	DB 35% - DC 27%- G 16% - D 14%	43%	49%	5%	3%
	PHTK>500	G 69%	47%	52%	1%	0,5%

Tableau 4: Test d'homogénéité des variances de Levene en groupant les accessions en fonction de leur classification raciale.

	Variance inter-race (ddl=4)	Variance intra-race (ddl=2495)	F	p
FLR	5692,91	148392,58	23,93	0,00
FLK	127719,91	406796,62	195,84	0,00
BT	28,03	783,63	22,31	0,00
PHTR	7722,65	2135571,63	2,26	0,06
PHTK	393797,15	4386020,19	56,00	0,00
PEDEX	907,61	66255,41	8,54	0,00
EHLG	2513,58	37572,75	41,73	0,00
EHWD	4845,42	20793,88	145,35	0,00
GRS	9,45	271,58	21,71	0,00
SWT	86,95	758,17	71,53	0,00

Tableau 5: Analyse du chi-2 testant les différences entre la distribution des accessions observée dans la collection, et celle attendue sous l'hypothèse d'indépendance statistique entre les races, les origines géographiques, et les caractères phénotypiques.

Variables	races	ddl	continents	ddl	latitudes	ddl	races x continents	ddl	races x latitudes	ddl
PIG	17,69	4	5,92 NS	4	2,34 NS	3	2248,58	16	1475,65	12
MCR	211,94	12	23,76	12	211,23	9	2286,22	48	1461,03	36
EHCS	1873,67	36	438,88	36	297,26	27	2327,22	144	1546,52	108
GLC	599,75	48	290,22	48	335,60	36	2423,64	192	1482,55	144
COV	3273,21	16	454,27	16	569,28	12	1984,33	64	1255,95	48
GRC	2513,00	48	419,75	48	548,08	36	2248,29	192	1430,17	144
COR	1513,31	16	65,47	16	346,53	12	2317,04	64	1451,39	48
LUS	1192,50	4	90,74	4	133,84	3	2260,65	16	1359,52	12
SC	796,60	4	51,24	4	37,07	3	2302,03	16	1519,04	12
Total	11991,68	220	1840,24	220	2481,23	141	20398,00	880	12981,84	660

NS: non significative (p>0,05)



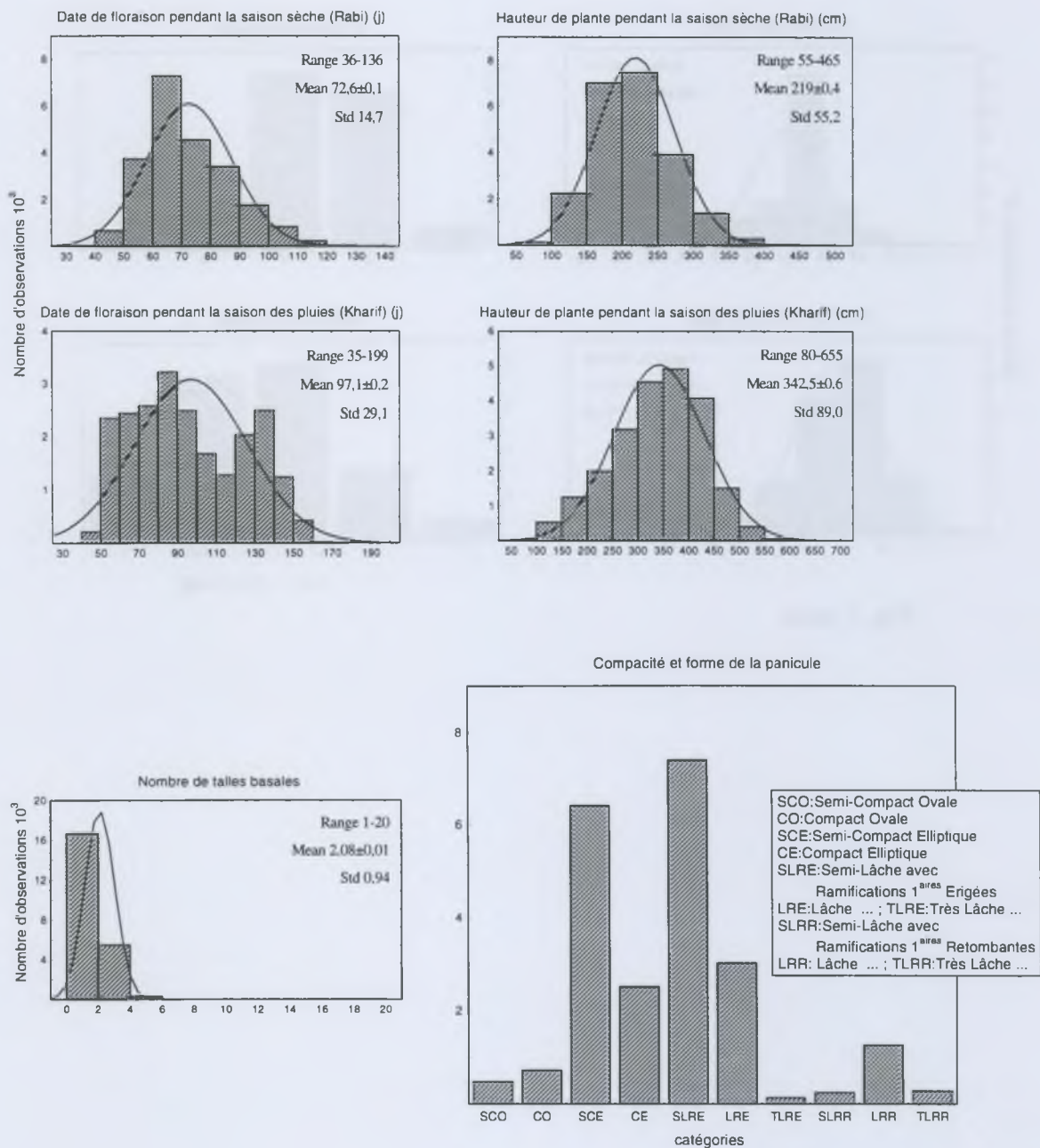


Fig. 1: Distribution des fréquences des sorghos cultivés pour les caractères continus et discrets.

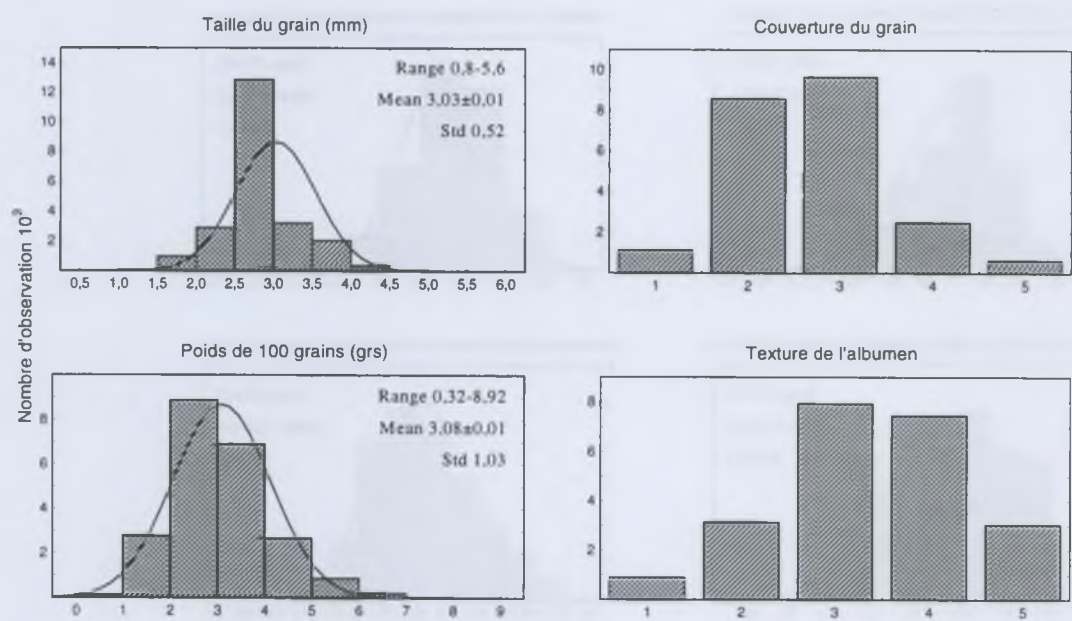


Fig. 1: suite

Core Collection of the Genetic Resources of Sorghum: I. Stratification  
Based on Eco-Geographical Data.

Grenier C, Bramel-Cox PJ, Hamon P

*Accepté sous réserve de corrections à Crop Sciences*

---



## CORE COLLECTION OF SORGHUM:

### I. STRATIFICATION BASED ON ECO-GEOGRAPHICAL DATA.

C. Grenier,\* P. J. Bramel-Cox, and P. Hamon

C. Grenier, CIRAD, BP 5035, 34 032 Montpellier Cedex 1 France; P. J. Bramel-Cox, GREP, ICRISAT Patancheru, 502 324 A.P., India; and P. Hamon, Université Montpellier III/IRD, BP5045, 34 032 Montpellier Cedex 1 France. Received\_\_\_\_\_.

\*Corresponding author (grenier@cirad.fr).

#### EXECUTIVE SUMMARY

A large collection of sorghum landraces [*Sorghum bicolor* (L.) Moench] is maintained at ICRISAT and the establishment of a core collection from random stratified sampling is considered that first required a hierarchical stratification of the collection. Grenier et al. proposed a way to stratify the collection using the flowering date and the plant height recorded during the two tropical seasons, where variations in day length allow differentiation between landraces for their response to photoperiod. Thus, collection was organized into four clusters that pertain to photoperiod sensitivity and to the specific adaptation within the collection. Moreover, great morpho-agronomic diversity within cluster was found. From the four clusters of photoperiod sensitivity, random samplings can be performed to establish core collections.

#### ABSTRACT

ICRISAT conserves a large (36 719 accessions) collection of sorghum, *Sorghum bicolor* (L.) Moench, in India. This collection of cultivated and wild sorghums was acquired over 25 years of collection and acquisition from 90 countries. A study of the morpho-agronomic diversity is used to describe the genetic structure of the collection. Morphological traits like date of flowering and plant height can be affected by day length variation. These two characters were highly correlated to the latitudinal and racial distributions of landraces. Thus, stratifying the entire collection for response to photoperiod, estimated by flowering date and plant height, was indicative of a major source of specific adaptation within the collection. This clustering resulted in four clusters, which described the sensitivity of the genotypes to the photoperiod in the latitudinal range where farmers selected them. These four clusters will be the basis for a random stratified sampling to establish core collections following different strategies.

Sorghum is the fifth most important cereal based on world production. It has a very long history as an essential component of the cropping systems and diets of subsistence farmers in the semi-arid tropics. In developing countries, its continued improvement depends upon the utilization of the genetic variability originally maintained in landraces by traditional agricultural practices.

Taxonomically, *Sorghum bicolor* (L.) Moench subs. *bicolor* has 15 races viz., the 5 basic races of bicolor, durra, caudatum, kafir, guinea, and their 10 intermediates (Harlan and de Wet, 1972). Numerous studies have shown that phenotypic divergence among and within landrace populations is related to geographical distance between countries of origin (Spagnoletti Zeuli and Qualset, 1987; Peeters and Martinelli, 1989; Schoen and Brown, 1993; Spagnoletti Zeuli and Qualset, 1993).

At present, 30 950 accessions of cultivated sorghum are conserved in the gene bank at ICRISAT Asia Center (IAC), Patancheru, India. These cultivated sorghum accessions have been compiled from 40 countries in Africa, 24 in Asia, 11 in Europe, 13 in the Americas plus Russia, and Australia. Sorghum landraces are known to respond differently to day length and temperature, according to their specific geographical adaptation. Temperature and photoperiod responses have been related to particular uses in traditional cropping systems (Chanterreau et al., 1998). Assessments of the impact of photoperiod on sorghum development have found sensitivity mainly within tropical genotypes (Quinby, 1966; Miller et al., 1968a; Miller et al., 1968b; Caddel and Weibel, 1971; Major et al., 1990; Vaksman et al., 1998). Accessions which have originated from different photothermal environments show variation in the growing degree-days required for panicle initiation due to differences in inherent earliness rather than responsiveness to temperature (Craufurd et al., 1998). Moreover, a genetic component in the floral initiation and differentiation was found by Miller et al. (1968b). These authors highlighted that during the long day season, flowering was under the control of a few major genes while under tropical conditions short day lengths from a different set of genes initiated it.

Farmers from West Africa preferentially sow highly adaptive landraces for the diversity of the environmental conditions (Kouressy et al., 1998). They actively select the landraces that will guarantee a maturity corresponding with the end of the rainy season, so that they can limit damage from birds and weather conditions (Trouche et al., 1998). Thus, sorghum photoperiod sensitivity has been subjected to selective pressure imposed both by nature and man.

Adaptation of sorghum to the diverse environments in which it is grown is determined largely by photoperiod sensitivity and minimum time to flower, the former one determines broad adaptation to latitude (day length) (Craufurd et al., 1999). For tropical areas, the year is divided into two major seasons defined by the monsoons, and days are longer during the rainy season than during the post-rainy season. There is no such variation in day length within a year at the equatorial latitudes. Consequently, introduced into tropical areas during the longest days of the rainy season, equatorial and tropical landraces adapted to short day length can delay flowering and even fail to flower, because of a non-optimal photoperiod. Those photoperiod sensitive landraces reach the flowering stage earlier during the post rainy season when days are shorter, and consequently have a shorter plant height. Therefore, photoperiod sensitivity could be defined by both the days to flowering and the plant height in the rainy and post-rainy season.

In order to study further the better way for the constitution of a set composed with a minimal number of accessions that represent at better the global diversity of the base collection, we are first interested in the analysis of the entire landrace collection. Thus, the



objective of this study was to assess the use of photoperiod sensitivity for hierarchical stratification of the entire landrace collection.

## MATERIALS AND METHODS

Initially, the entire ICRISAT sorghum collection was reduced to accessions corresponding to landraces from a latitude range of 40°N to 40°S with complete passport information and characterization data. Beyond this geographical range, few accessions were expected to be original traditional cultivars. This landrace collection consisted of 22 473 landraces compiled from 76 countries, which represented 62% of the collection conserved at ICRISAT. The difficulty encountered in establishing geographical grouping was resolved by stratifying the latitudes into four ranges of 20° each (i.e., latitude range 1 (LR-1) = 10°S - 10°N, latitude range 2 (LR-2) = 20°S - 10°S, 10°N - 20°N, latitude range 3 (LR-3) = 30°S - 20°S, 20°N - 30°N, latitude range 4 (LR-4) = 40°S - 30°S, 30°N - 40°N). A crosstabulation table was constructed that placed the landraces into 60 groups according to their race classification (Table 1, row totals) and latitudinal range (Table 1, column totals).

For characterization, the accessions were grown on vertisol soils at Patancheru, India (17°25'N, and 78°E). Field observations were made over several years from 1975 to 1996 in both the kharif (rainy, relatively higher mean temperature 27°C, long day) and the rabi (post-rainy, lower mean temperature 22°C, short day) seasons. Kharif is typically from June to December, while rabi usually extends from September to April. These two different environmental conditions are typical of the semi-arid tropics (Appa Rao et al., 1996). Data were recorded for 10 quantitative characters as described in Appa Rao et al. (1996). Days to 50% flowering were recorded as the mean emergence date to the date when 50% plants had started flowering (i.e., ca. 15 out of the 30 plants per row per accession). Plant height was observed at maturity and recorded as the length of the main stalk (cm) for ten plants. Both of these characters were measured in the two seasons, kharif and rabi. Peduncle exertion (cm), panicle length (cm), panicle width (cm) and number of basal tillers were measured on ten plants in the field. After harvest, grain size (mm) and one hundred seed weight (g) were measured. Photoperiod response was defined by the two traits related to the vegetative development, namely number of days to 50% flowering and plant height. The relative variables (FL and PHT) were defined by the differences in measurements obtained during the long day length of the rainy season (kharif) and the shorter day length of the post-rainy season (rabi). FL was calculated as the difference in days to flowering in kharif (FLK) and rabi (FLR). PHT was calculated as the difference in plant height in kharif (PHTK) and rabi (PHTR).

Before using the information on morpho-agronomic characters available in the data base, a validation of the data recorded from the 21 years of characterization held at IAC was performed with 642 accessions selected to represent a broad range of the germplasm collection (the 15 races and 43 countries). During two complete years from 1996 to 1998, field trials were done for these 642 accessions plus a check (IS 1054) repeated within each field and between seasons and years. For quantitative characters, the records from 1996 to 1998 showed positive and significant ( $p < 0.05$ ) correlation with those found in the data base (data not shown). Thus, from this original data file on 22 473 landraces, FL and PHT were calculated and split into four classes of the same size according to their quartiles (Table 2). Thus, a disjunctive table was obtained with the 22 473 accessions and the eight variables (four variables for FL and four variables for PHT). Then, the frequency distribution within the 60 race-latitudinal groups for each of the eight variables was calculated. In our study, we have chosen the K-means procedure following a Ward's clustering in order to well define the



clusters and their constitution. Indeed, the K-means clustering method requires an *a priori* number of clusters prior to classify the accessions into clusters. Thus, a Ward's clustering was performed on variables related to the photoperiod reaction in order i) to provide a number of clusters that well stratified the entire landrace collection and ii) to give a biological sense to such structure. Thus, in a first step, the matrix of the 60 groups by eight variables, were subjected to a cluster analysis using Ward's linking procedure based on the Euclidean distance between clusters (StatSoft, 1997). In a second step, the K-means procedure, compared to an "ANOVA in reverse" which minimizes within-cluster variability while maximizing variability between clusters, classifies at better each group in the given clusters.

In addition, the diversity of the entire landrace collection was assessed by a principal component analysis of the standardized data of the ten quantitative characters. Scores were calculated for individual landraces from the first two components. Fifty landraces were taken at random within each of the four clusters. These were plotted for their first two principal component scores.

## RESULTS AND DISCUSSION

The classification of the landrace collection into four latitudinal origins (Table 1) found the greatest number (86.8%) from equatorial and tropical areas, i.e., LR-1 and LR-2. The classification of the landrace collection into the 15 taxonomic races found that the highest proportion were caudatum, durra, guinea, durra-caudatum and guinea-caudatum forms, which represented 82.2% of the total landraces in this collection.

The distribution of the landraces within the latitudinal classes and the four frequency classes of FL and PHT (as described in Table 2) is shown in Fig. 1. For the lower latitudes (LR-1, i.e., equatorial zones), the distribution of the accessions was skewed toward the highly photoperiod responsive class for both variables. Around half of the landraces from the total collection belong to the LR-2 (47.1%) and were evenly distributed over the four frequency classes for both variables. Conversely to the equatorial latitudes, a higher frequency of accessions from the temperate zones (LR-3 and LR-4) belong to FL1 and PHT1 where 25% of the total collection was insensitive to photoperiod.

Similar results were found when the distribution of FL and PHT within the 60 groups (Fig. 2) was examined. These graphs illustrate the relationship between racial classification, geographical adaptation and photoperiod response. Generally, the frequency of accessions insensitive to photoperiod increased with the latitudinal class. However within the same latitudinal class, races reacted differently to the day length variation. The frequency distribution of guinea, durra, and their intermediate forms within any latitudinal class was skewed toward the most sensitive photoperiod response, even at the temperate latitude. For the temperate latitude (LR-4), only 10% of the landraces show sensitivity to day length variation and 47% of these accessions belong to the guinea race. Those guinea landraces from temperate zones represent 4% of all the guinea accessions held in the landrace collection. This result is in agreement with other studies defining West African guinea landraces as plants with short-day photoperiod reaction responding to day length by hastening their development (Chanterreau et al., 1998; Trouche et al., 1998; Vaksman et al., 1998).

Conversely, for the kafir race more than 90% of the accessions were photoperiod insensitive (FL1 and PHT1). Only 1.2% of the kafir accessions was classified as moderately responsive to photoperiod at the latitudes where 82.6% of the landraces were highly responsive (LR-1). Moreover, 40% of the moderately photoperiod responsive accessions at the equatorial latitude were race kafir. Other studies have shown the same photoperiod insensitivity of accessions from the kafir race that have originated from high latitudes (Chanterreau et al., 1998). These observations of the distribution of the race-latitudinal groups

into different classes for the photoperiod response agree with the fact that photoperiod sensitivity is due to the complex interactions between environmental and genetic components.

With the K-means clustering, 95% of the race-latitudinal groups were well classified in comparison with the Ward's clustering. Thus, the partition of the 60 groups into four clusters can be validated. The constitution of each cluster is given in Table 3. In general, each cluster was associated with a particular latitudinal class. Landraces that originated from LR-3 and LR-4 mainly belong to the clusters that include the mildly photoperiod sensitive and photoperiod insensitive groups, respectively. The third cluster, mostly composed of groups from tropical latitudes (LR-2), was designated as photoperiod sensitive. The fourth cluster of highly photoperiod sensitive groups consisted of landraces mainly from equatorial latitudes (LR-1).

Race groups from the same latitude mainly clustered together. Nevertheless, some exceptions were noted. Equatorial groups that contained landraces that were known to be highly photoperiod sensitive were clustered with photoperiod insensitive temperate landraces. This was the case for accessions from the kafir race and its intermediate forms and was consistent with the results found in Fig 2. As an example, the guinea-kafir intermediate from equatorial latitude (1GK) was clustered with the kafir race from temperate latitude (4K) (Table 3). The reduced photoperiod sensitivity of the kafir forms from equatorial latitude and the exclusion of the equatorial durra and caudatum races from the highly photoperiod sensitive cluster cannot be explained solely by adaptation to specific environments through natural selection. Possible explanations could be that it reflects farmers' specific selection for a particular use or cropping system, or that it results from the effects of migration of landraces from latitudes where they are adapted to different photoperiod. Another example of the effect of farmers' selection could be found in the highly photoperiod sensitive cluster where all the groups were from equatorial latitudes except two groups of guinea forms from tropical latitudes. The extent of photoperiod adaptation for each race can also be a factor explaining such conservative response to photoperiod.

Within this landrace collection, 16 race-latitudinal groups (1160 accessions) were classified as photoperiod insensitive. A second cluster with 15 race-latitudinal groups (1062 accessions) was classified as mildly photoperiod sensitive. The third cluster (10 630 accessions) included 19 race-latitudinal groups of photoperiod sensitive landraces. Finally, the fourth cluster (9621 accessions) consisted of 10 race-latitudinal groups of highly photoperiod sensitive landraces. Thus, the distribution of the accessions in the ICRISAT sorghum landrace collection according to photoperiod sensitivity is consistent with the evolution of the crop and the main use of landraces in cropping systems of the semi-arid tropics.

The factor analysis of the 10 quantitative morpho-agronomic characters explained 48.8% of the variance with the first two axes. The first axis accounted for 29.6% of the variance and the principal component scores were positively influenced by days to flower and plant height in the rainy season (factor loading 0.84 and 0.85, respectively) and in the post-rainy season (loading 0.70 for both variables). The second axis explained 19.2% of the variance and grain size (loading 0.85) and hundred seed weight (loading 0.80) influenced scores. Fifty landraces were randomly chosen within each cluster and the first two principal component scores were plotted (Fig. 3). Although four well-differentiated and non-overlapping clusters were obtained from the race-latitudinal groups, great intra-cluster diversity was found. The photoperiod sensitive and highly photoperiod sensitive landraces had a wider distribution of the morpho-agronomic diversity, as described by the first two component scores. There was variability for the characters that described the first two components, although the distribution of the accessions within the photoperiod insensitive



and mildly photoperiod sensitive clusters were reduced. For example, there were accessions within these two clusters that flowered early and had a shorter plant height with both small and large grains. Conversely, large seeded accessions had a smaller range of days to flower and plant height. Thus, there was no evidence of a sharply reduced diversity within the two clusters that would have the greatest value for breeding in the temperate regions. This result meets the objectives of the study since a stratification of the entire landrace collection was found that keeps a great morpho-agronomic diversity within cluster. From this point, random samplings could be performed within each cluster.

In conclusion, with the analysis of the photoperiod response, race classification and latitudinal origin, we found a clear structure to the diversity within this landrace collection of sorghum maintained at ICRISAT. For the core collection establishment, stratification according to the photoperiod sensitivity is efficient for the sampling strategies more especially as the clusters are of unequal size. Here the ratio is around 1/10. Such unbalanced representation of the clusters will allow avoiding the loose or the defective sampling of the small clusters. The use of a hierarchical stratification based upon an estimate of the photoperiod sensitivity simplified the diversity among the entire landrace collection into four well-defined clusters. There was evidence that both natural selection for adaptation to the environment and farmers' selection for specific use or specific cropping systems accounted for the morpho-agronomic diversity. This will enhance utilization of the landrace collection by allowing breeders to concentrate their evaluation on variation within the clusters as an initial focus. This will also assist in the definition of a core collection that can be used to enhance the maintenance, evaluation and utilization of the collection by breeders and farmers.

#### ACKNOWLEDGMENTS

We thank M. Deu and S. Chandra for their helpful remarks and review of the manuscript, J. Chantereau for his valuable comments and explanations on the sorghum physiology, and ICRISAT-GRD staff for their help for fieldwork and access to the database.

#### REFERENCES

- Appa Rao, S., K.E. Prasada Rao, M.H. Mengesha, and V. Gopal Reddy. 1996. Morphological diversity in sorghum germplasm from India. *Gen. Res. Crop Evol.* 43:559-567.
- Caddel, J.L., and D.E. Weibel. 1971. Effect of photoperiod and temperature on the development of sorghum. *Agron. J.* 63:799-802.
- Chantereau, J., M. Vaksman, I. Bahmani, M.A.G. Hamada, M. Chartier, and R. Bonhomme. 1998. Characterization of different temperature and photoperiod responses in African sorghum cultivars. *Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre.* p. 29-35. *In* A. Ratnadass et al. (ed.) *Actes de l'atelier de restitution du programme conjoint sur le sorgho ICRISAT-CIRAD, Bamako, Mali, Collection colloques, Montpellier, CIRAD-CA.*
- Craufurd, P.Q., A. Qi, R.H. Ellis, R.J. Summerfield, E.H. Roberts, and V. Mahalakshmi. 1998. Effect of temperature on time to panicle initiation and leaf appearance in sorghum. *Crop Sci.* 38:942-947.
- Craufurd, P.Q., V. Mahalakshmi, F.R. Bidinger, S.Z. Mukuru, J. Chantereau, P.A. Omanga, A. Qi, E.H. Roberts, R.J. Summerfield, and G.L. Hammer. 1999. Adaptation of sorghum: characterization of genotypic flowering responses to temperature and photoperiod. *Theor Appl Genet* 99:900-911.
- Harlan, J.R., and J.M.J. de Wet. 1972. A simplified classification of cultivated sorghum. *Crop Sci.* 12:172-176.
- Kouressy, M., M. Ouattara, and M. Vaksman. 1998. Importance du photopériodisme chez les sorghos tropicaux, conséquences pour un programme de sélection. p. 49-54. *In* A. Ratnadass et al. (ed.) *Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre. Actes de l'atelier de restitution du programme conjoint sur le sorgho ICRISAT-CIRAD, Bamako, Mali, Collection colloques, Montpellier, CIRAD-CA.*
- Major, D.J., S.B. Rood, and F.R. Miller. 1990. Temperature and photoperiod effects mediated by the sorghum maturity genes. *Crop Sci.* 30:305-310.
- Miller, F.R., D.K. Barnes, and H.J. Cruzado. 1968a. Effect of tropical photoperiods on the growth of sorghum when grown in 12 monthly plantings. *Crop Sci.* 8:499-502.



- Miller, F.R., J.R. Quinby, and H.J. Cruzado. 1968b. Expression of known genes of sorghum in temperate and tropical environments. *Crop Sci.* 8:675-677.
- Peeters, J.P., and J.A. Matrinelli. 1989. Hierarchical cluster analysis as a tool to manage variation in germplasm collections. *Theor. Appl. Genet.* 78:42-48.
- Quinby, J.R. 1966. Fourth maturity locus in sorghum. *Crop Sci.* 6:516-518.
- Schoen, D.J., and A.H.D. Brown. 1993. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proc. Natl. Acad. Sci. (USA)* 90:10623-10627.
- Spagnoletti Zeuli, P.L., and C.O. Qualset. 1987. Geographical diversity for quantitative spike characters in world collection of durum wheat. *Crop Sci.* 27:235-241.
- Spagnoletti Zeuli, P.L., and C.O. Qualset. 1993. Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. *Theor Appl Genet* 87:295-304.
- StatSoft. 1997. Statistica for Windows. USA, StatSoft, Tulsa, USA.
- Trouche, G., M. Vaksman, J. Chantereau, M. Kouressy, H.D. Maiga, and C. Barro. 1998. Etude du déterminisme génétique du photopériodisme des sorghos guinea. p. 37-47. *In* A. Ratnadass et al. (ed.) Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre. Actes de l'atelier de restitution du programme conjoint sur le sorgho ICRISAT-CIRAD, Bamako, Mali, Collection colloques, Montpellier, CIRAD-CA.
- Vaksman, M., J. Chantereau, I. Bahmani, M.A.G. Hamada, M. Chartier, and R. Bonhomme. 1998. Influence of night temperature on photoperiod response of a West African guinea sorghum landrace. p. 23-28. *In* A. Ratnadass et al. (ed.) Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre. Actes de l'atelier de restitution du programme conjoint sur le sorgho ICRISAT-CIRAD, Bamako, Mali, Collection colloques, Montpellier, CIRAD-CA.

## FIGURE LEGENDS

Figure 1. Frequency distribution according to the latitudinal ranges (LR) for the two variables relevant to the photoperiod sensitivity in the ICRISAT sorghum landrace collection: difference between the two seasons in (A) flowering date (FL), and (B) plant height (PHT). Each variable was split into 4 classes according to their quartiles (Table 1). Each variable representing 25% of the total collection characterizes a photoperiod sensitive class.

Figure 2. Photoperiod sensitivity distribution of frequencies for days to flowering (FL) or plant height (PHT) classes according to the latitudinal range and the racial classification for the ICRISAT sorghum landrace collection. Races are designated by the first letter name, i.e., B for bicolor, C for caudatum, D for durra, G for guinea, K for kafir and for example CB for caudatum-bicolor. Number preceding the letter race name indicates the latitudinal range, i.e., 1: LR-1, 2: LR-2, 3: LR-3, and 4: LR-4. FL1 and PHT1 are represented with lozenges, FL2 and PHT2 with blank, FL3 and PHT3 with dashes, FL4 and PHT4 with pattern of squares.

Figure 3. Principal component analysis of diversity for ten quantitative morpho-agronomic characters in the ICRISAT sorghum landrace collection. Plot of the rotated (Varimax) factor-scores for 50 accessions randomly selected within each cluster of photoperiod sensitivity. Cluster membership was defined from the eight photoperiod variables using K-means procedure.

Table 1. Crosstabulation table of the sorghum landrace collection based upon racial classification (rows) and latitudinal classification (columns) with the percentage of landraces belonging to each cell.

Races	Latitudinal Range				All
	LR-1: 10°S-10°N	LR-2: 20°S-10°S, 10°N-20°N	LR-3: 30°S-20°S, 20°N-30°N	LR-4: 40°S-30°S, 30°N-40°N	
B: bicolor	0.9	1.2	0.3	0.8	3.2
C: caudatum	14.6	5.6	0.8	0.8	21.8
D: durra	5.9	11.7	2.3	0.3	20.2
G: guinea	7.0	9.7	0.4	0.2	17.4
K: kafir	0.0	0.3	0.9	1.1	2.3
CB: caudatum-bicolor	0.9	1.6	0.4	0.5	3.4
DB: durra-bicolor	2.8	2.2	0.5	0.1	5.6
GB: guinea-bicolor	0.6	0.3	0.0	0.0	1.0
KB: kafir-bicolor	0.0	0.1	0.1	0.0	0.2
DC: durra-caudatum	3.0	8.3	0.6	0.3	12.1
GC: guinea-caudatum	3.7	5.3	1.1	0.5	10.7
KC: kafir-caudatum	0.0	0.2	0.4	0.2	0.9
GD: guinea-durra	0.1	0.4	0.1	0.0	0.6
KD: kafir-durra	0.0	0.2	0.2	0.1	0.5
GK: guinea-kafir	0.0	0.0	0.1	0.0	0.1
Total	39.7	47.1	8.2	5.0	100.0



Table 2. Definition of the variables obtained from the difference in measurements between rainy (kharif) and post-rainy (rabi) season in flowering date (FL) (days), and in plant height (PHT) (cm) for the sorghum landrace collection. Q1, Q2, Q3 are the first, second and third quartiles.

Ranges	FL = FLK - FLR	PHT = PHTK - PHTR
[min - Q1]	[-58, 4] ≡ FL1	[-170, 65] ≡ PHT1
[Q1 - Q2]	[4, 23] ≡ FL2	[65, 130] ≡ PHT2
[Q2 - Q3]	[23, 43] ≡ FL3	[130, 180] ≡ PHT3
[Q3 - max]	[43, 139] ≡ FL4	[180, 425] ≡ PHT4

Table 3: Constitution of each cluster of photoperiod reaction defined from the K-means clustering performed on the sorghum landrace collection.

Insensitive	Mid sensitive	Sensitive	Highly sensitive
1GK	1K	1D	1B
1KD	1KC	1CB	1C
2GK	2KB	1KB	1G
2K	3B	2B	1DB
3C	3DC	2C	1GB
3K	3GC	2D	1DC
3CB	3GD	2CB	1GC
3KB	4B	2DB	1GD
3KC	4C	2DC	2G
3KD	4DB	2GC	2GB
3GK	4GB	2GD	
4K	4DC	2KC	
4CB	4GC	2KD	
4KB	4GD	3D	
4KC	4GK	3G	
4KD		3DB	
		3GB	
		4D	
		4G	

Figure 1

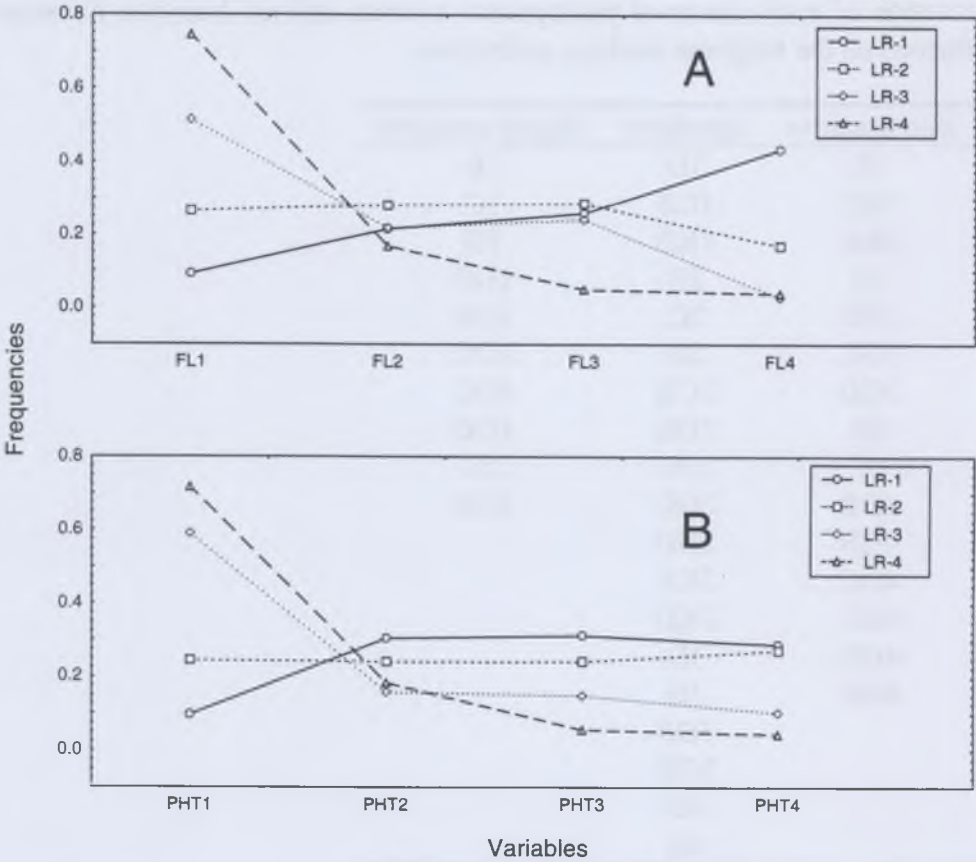
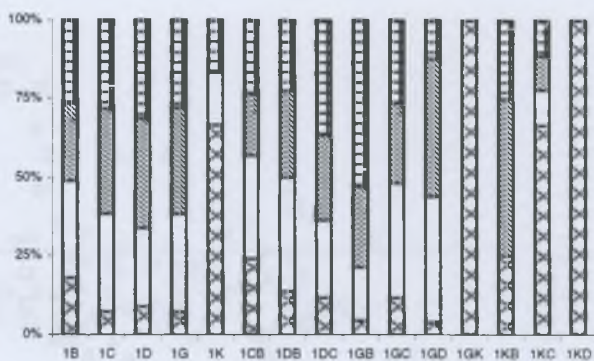
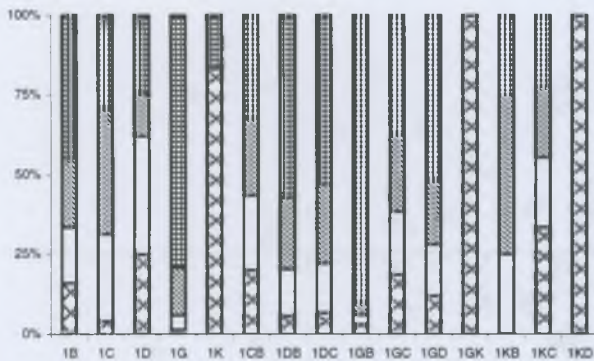




Figure 2

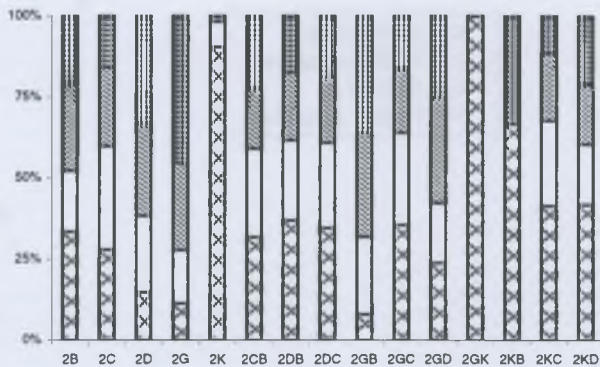
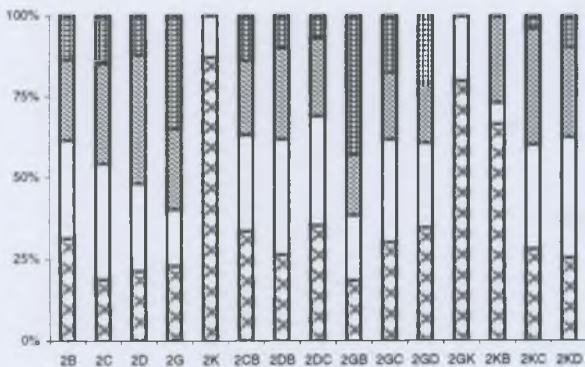
FL

PHT



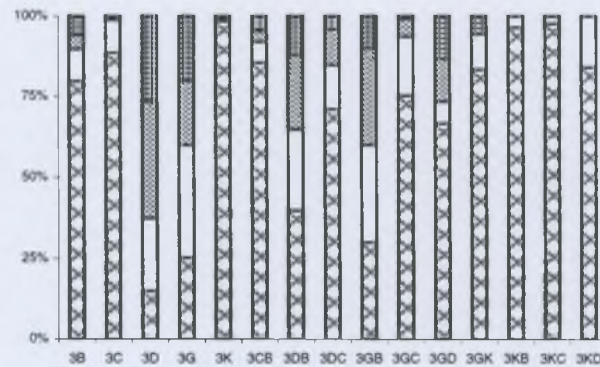
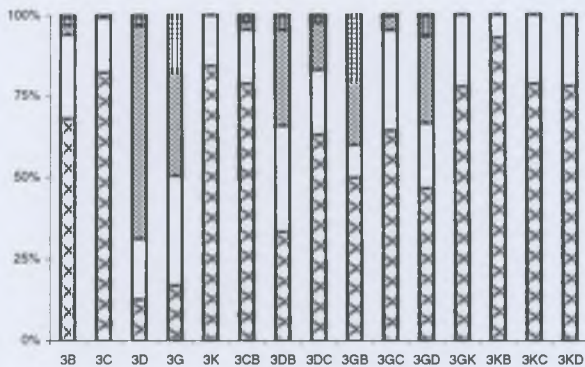
FL

PHT



FL

PHT



FL

PHT

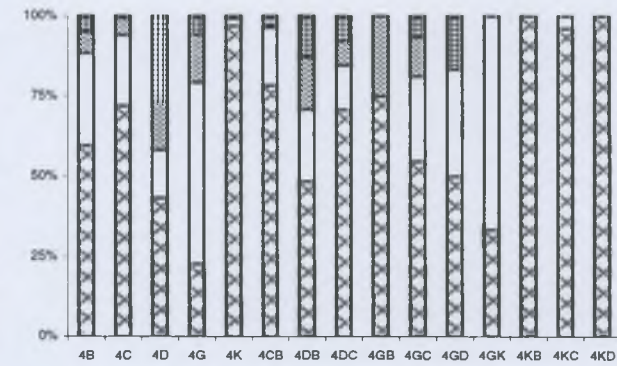
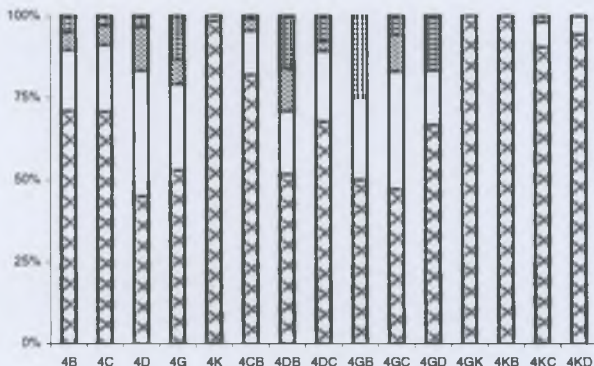
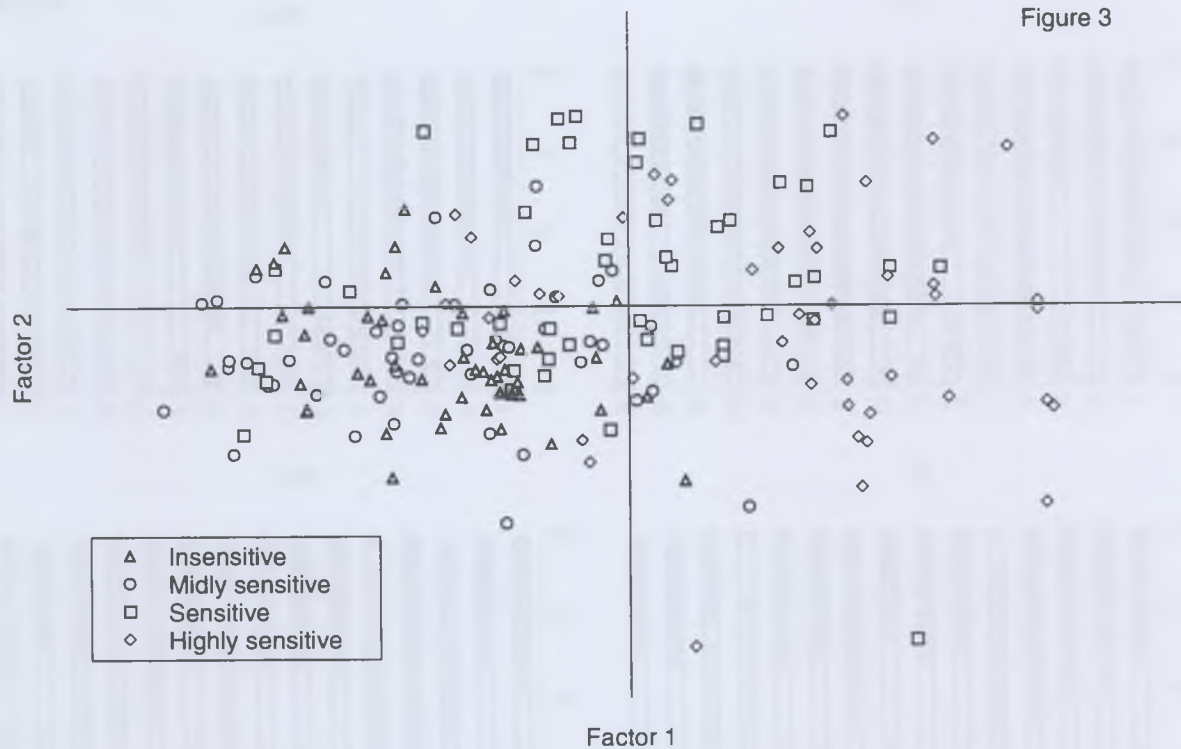


Figure 3



Grenier C, Hamon P, Bramel-Cox PJ, 1999. Core Collection of the Genetic Resources of Sorghum: II. Comparison of Three Random Sampling Procedures.

*Accepté sous réserve de corrections à Crop Sciences*

---



## CORE COLLECTION OF SORGHUM:

### II. COMPARISON OF THREE RANDOM SAMPLING STRATEGIES.

C. Grenier,\* P. Hamon, and P. J. Bramel-Cox

C. Grenier, CIRAD, BP 5035, 34 032 Montpellier Cedex 1 France; P. Hamon, Université Montpellier III/IRD, BP5045, 34 032 Montpellier Cedex 1 France; and P. J. Bramel-Cox, ICRISAT-GREP Patancheru, 502 324 A.P., India. Received\_\_\_\_\_.

\*Corresponding author (grenier@cirad.fr).

Abbreviations: FL = difference in days to flowering (days) between the rainy and the dry seasons - PHT = difference in plant height (cm) between the rainy and the dry seasons.

#### EXECUTIVE SUMMARY

The large sorghum landrace collection [*Sorghum bicolor* (L.) Moench] maintained at ICRISAT, and organized into four clusters of photoperiod sensitivity, was subject to sampling in order to propose core collections. Grenier et al. performed different random stratified sampling strategies and core collection were compared to each other and to the landrace collection for the distribution of morpho-agronomic diversity and geographical origins. Core collection established from a proportional sampling shown similar pattern of distribution than the landrace collection. Conversely, logarithmic sampling differed and better represents landraces belonging to the smaller clusters and thus a larger range of photoperiod sensitivity. The logarithmic sampling strategy can be chosen as a way to establish a core collection with the widest range of adaptation to photoperiod.

#### ABSTRACT

Since 1972, ICRISAT has maintained a large collection of sorghum, *Sorghum bicolor* (L.) Moench subs. *bicolor*, in India. The collection size has continuously increased and the total number of accessions at present conserved in the gene bank has reached about 36 000 accessions. The need to help management was considered and this study was conducted to establish core collections. This sorghum collection was earlier stratified into four clusters according to the photoperiod sensitivity. Then, considering the core collection strategy, we used three random sampling procedures to determine the specific accessions to be included in the core; i.e. a constant portion (C), a proportional (P), and a proportional to the logarithm (L) of the photoperiod group's size sampling strategy. Both the core C and L were significantly different from the landrace collection with better representation of the smallest groups, such as landraces insensitive to photoperiod. Despite differences between the three core collections, estimates of global diversity through the Shannon-Weaver Diversity Indices were of the same magnitude as the landrace collection. When compared, the core C and L were significantly different. Core L sampled better for the characters, the race, and the latitudinal classes that were related to the photoperiod sensitive landraces. Thus for establishing a core collection with the widest range of adaptation to photoperiod, we propose to use a logarithmic sampling strategy, which identifies a broadly adapted set of genotypes.

A large collection, 36 719 accessions, of sorghum is conserved at ICRISAT. This collection is made up of breeding lines (14.5%), wild species (1.1%), and traditional varieties i.e. landraces (84.3%). The management of this collection involves acquisition of new accessions, conservation, characterization, evaluation, enhancement and distribution of these germplasm (Brown, 1995). To facilitate the management and the use of genetic resources the concept of 'core collection' was proposed by Frankel 1984 (in Brown, 1989a) to provide efficient access to the whole collection (Brown, 1989a). Thus, a core collection was defined as a limited set of accessions chosen to represent the genetic spectrum in the whole collection. Under the sampling theory of selectively neutral alleles, Brown (1989b) suggested a core collection size of about 10% of the entire collection.

To establish this core collection, a number of random sampling strategies were proposed by Brown (1989b). The initial step would be to stratify a collection. Then each stratum could be sampled by either a constant number for all the clusters (C strategy), a number proportional to the cluster size (P strategy), or a number proportional to the logarithm of the cluster size (L strategy). The P strategy has been the most commonly used to constitute a core collection (Holbrook et al., 1993; Basigalup et al., 1995; Prasada Rao and Ramanatha Rao, 1995; Ortiz et al., 1998). The C strategy has been used only by Vaughan (1991) on the IRRI rice collection, and the L strategy was used by Iguarta et al. (1998) on the Spanish barley collection. From these studies, two questions have arisen on the use of these three possible sampling strategies. First, do the different strategies retain an equivalent level of diversity from the original collection? Second, how do the core collections identified from these various sampling strategies compare?

Previously, the world sorghum collection was stratified into four clusters according to photoperiod sensitivity (Grenier et al., submitted). The landrace collection held at ICRISAT was considered as 60 groups obtained from the crosstabulation table between the racial membership of the landraces (15 races) and their geographical origin (four latitudinal ranges). For flowering date and plant height, differences in the records between the two seasons of the semi arid-tropic (the rabi with shorter day length, and the kharif with longer day length) gave two continuous variables. Each one was translated into qualitative variable by splitting into four frequency classes using the quartiles. Then, from the frequency distribution of landraces within race-latitudinal groups (60 groups) into the eight variables (four variables for flowering date and four variables for plant height) a K-means clustering procedure was used to get four clusters that pertain to the reaction to photoperiod. Thus, the four clusters obtained described the adaptations of the accessions due to both natural selection and farmers' selection. In this study, we established three core collections from the three random sampling strategies applied to this stratified landrace collection. The objectives of this study were 1) to compare the three core collections with each other and to the original landrace collection for their representation of diversity, and 2) to propose a better way to sample the diversity of the entire landrace collection.

## MATERIAL AND METHODS

From the entire collection of cultivated sorghum maintained at ICRISAT, only landraces assembled from the range of latitude of 40°N to 40°S with complete data for all the characteristics (passport, qualitative and quantitative) were considered. This collection represents 22 473 accessions.

For characterization, the accessions were grown on vertisol soils at Patancheru, India (17°25'N and 78°E). Field observations were made over several years from 1975 to 1996 in both the rainy (kharif) and the post-rainy (rabi) seasons. These two different environmental conditions are typical of the semi-arid tropics (Appa Rao et al., 1996). The rainy season



occurs during long days (June to September) while the dry season occurs during short days (September to April). Data were recorded for ten quantitative characters (Table 1) as described in Appa Rao et al. (1996). Days to 50% flowering were recorded as the mean emergence date to the date when 50% plants had started flowering (i.e., ca. 15 out of the 30 plants per row per accession). Plant height was observed at maturity and recorded as the length of the main stalk (cm) for ten plants. Both of these characters were measured in the dry and rainy seasons. Peduncle exertion (cm), panicle length (cm), panicle width (cm) and number of basal tillers were measured on ten plants in the field. After harvest, grain size (mm) and one hundred seed weight (g) were measured. Two relative variables, FL and PHT, were defined as the differences in measurements obtained during the long day length of the rainy season and the shorter day length of the dry season. FL was calculated as the difference in days to flowering and PHT as the difference in plant height in the rainy and dry seasons. The nine qualitative characters (Table 2) recorded were leaf and midrib pigmentation, shape of the inflorescence, glume and grain color, glume covering, type of grain (corneous and lustrous), and presence or absence of subcoat.

The entire collection was stratified into four clusters by a K-means procedure (Statistica, 1997) performed on a matrix of 60 groups and 8 variables. The 60 groups were defined by the racial membership of the landraces for the 15 races defined according to the Harlan and de Wet racial classification (Harlan and de Wet, 1972) and the geographical origin defined with four latitudinal ranges (Grenier et al., submitted). The eight variables were obtained from the two photoperiod response variables, FL and PHT, split into four frequency classes according to the quartiles. From this matrix, K-means clustering procedure was performed using an *a priori* number of four clusters. From this stratification, the first cluster consisted of 1160 accessions classified as photoperiod insensitive. The second cluster included 1062 accessions classified as mildly photoperiod sensitive. The third cluster consisted of 10 630 accessions classified as photoperiod sensitive landraces. Finally, the fourth cluster consisted of 9621 accessions classified as highly photoperiod sensitive landraces (Grenier et al., submitted).

Core collections that represented ten percent of the landrace collection (2247 accessions) were established from three random sampling procedures. The C strategy sampled at random a constant number (562) of accessions from each cluster irrespective of its size. The P strategy sampled at random ten percent of the number of accessions within each cluster (i.e., 116, 106, 1063, and 962 accessions). The L strategy sampled at random proportionally from the logarithm of the number of accessions within each cluster (i.e., 488, 482, 642, and 635 accessions).

For the ten quantitative characters, comparisons between the core collections and the landrace collection were based upon Levene's test of homogeneity of variance as well as Newman and Keuls' test for post-hoc comparison of means performed only when variances were comparable (Statsoft, 1997).

Because of the categorical nature of the qualitative characters, the passport data and the racial membership, a non-parametric statistical procedure was used to evaluate and compare the distributions. The chi-square test was used to compare the distribution of accessions for the nine qualitative characters, passport data and racial distribution in the landrace collection to those of the core collections. In the last case, the chi-square test was obtained from the racial membership determined with the Harlan and de Wet's race classification. Passport data were compared with the frequency distributions for the geographical origin i.e. continent and latitudinal ranges. The distributions observed in each core were compared to the distribution expected in the landrace collection adjusted for accession number (10%). The chi-square test was also used to compare the distributions of



the three core collections. Distributions for races and for passport data within each core collection were considered.

For each core, global diversity was estimated by the Shannon-Weaver Diversity Index as presented by Poole (1974; in Li et al., 1996). This phenotypic diversity index is based upon the frequency distributions for the nine qualitative characters. This index also included the ten quantitative characters transformed into four phenotypic classes defined by the four quartiles of the landrace collection. The Shannon-Weaver Diversity Index,  $H_C'$  was estimated using:

$$H_C' = - \sum_{i=1}^n p_i \log_e p_i$$

where for a given character C,  $n$  was the number of phenotypic classes (for the qualitative characters  $n = 2$  to 13 descriptor states, and for quantitative characters  $n = 4$  frequency classes), and  $p_i$  was the proportion of the total number of entries in the  $i^{\text{th}}$  class. Due to its additive property (Poole, 1974 in Li et al., 1996), Shannon-Weaver Diversity Indices obtained for each character were pooled for each core collection and landrace collection. Then, means and standard errors were calculated from the non-standardized indices, and the variance was approximated by

$$\text{Var}(H') = [\sum p_i \log_e^2 p_i - (\sum p_i \log_e p_i)^2]/N + (n-1)/2N^2$$

where  $N$  was the number of observations (Hutcheson, 1970). The  $H'$  indices were compared by a t-test:

$$T = (H'_1 - H'_2)/[\text{Var}(H'_1) + \text{Var}(H'_2)]^{1/2},$$

with degrees of freedom:

$$\text{d.f.} = [\text{Var}(H_1) + \text{Var}(H_2)]^2 / [\text{Var}(H_1)^2/N_1 + \text{Var}(H_2)^2/N_2],$$

where  $N_1$  and  $N_2$  were the numbers of entries in the two groups.

To compare indices obtained for different characters, a standardized index  $\text{SDI}_C$  (range value from 0 to 1) was calculated as:

$$\text{SDI}_C = H_C' / \log_e n$$

## RESULTS

Considering the ten quantitative morphological characters, comparisons between the core collections and the landrace collection were based upon tests for homogeneity of variances and, when possible, means comparison (Table 1). The results indicated that for core collection P, none of the ten traits had a significantly different variance from the landrace collection. Furthermore, means were not significantly different from those of the landrace collection. For both the core C and the core L, homogeneous variances were found for two traits, days to flowering in kharif and panicle width. However, compared to the landrace collection, only core C did not have a significantly different mean for panicle width.

The chi-squares compared observed (in the core collection) versus expected (obtained for 10% of the landrace collection) distributions calculated for the nine morphological qualitative characters (Table 2). For all these characters in the core collection P, the distribution was not significantly different from the observed in the landrace collection. Conversely, for nearly all characters, core C and L were significantly different from the distributions expected in the landrace collection. Core C and core L exhibited similar frequencies for leaf pigmentation, and only the core L did not differ from the landrace collection for grain lustre. Thus, when the qualitative and quantitative characters were considered, the core collection P represented a similar pattern of diversity as the landrace collection while core C and core L exhibited different distributions.

The Shannon-Weaver Diversity Indices for the landrace collection and the three core collections, with mean and standard error were calculated from the nine qualitative characters

and the ten quantitative characters. The variance of the Shannon-Weaver Diversity Indices was used to compare these diversity estimates. Statistically similar morpho-agronomic diversity levels were obtained in the comparison of the landrace collection and the core collections ( $H' = 1.261 \pm 0.111$  for the landrace collection and  $1.258 \pm 0.111$ ,  $1.237 \pm 0.106$ , and  $1.231 \pm 0.106$ , for the core collection P, L and C, respectively). Based upon the standardized Shannon-Weaver Diversity Indices, relatively high levels of diversity were found ( $SDI = 0.853 \pm 0.053$ ,  $0.851 \pm 0.052$ ,  $0.840 \pm 0.050$ , and  $0.835 \pm 0.050$ , for landrace collection, core P, L and C, respectively). While, core C and core L had a different pattern of distribution from the landrace collection for most of the morpho-agronomic characters, the overall diversity was maintained in relation to the landrace collection.

The diversity of core collections was also compared for their deviation from the original distribution of accessions for the continent of origin, the latitudinal class and the race classification through a chi-square test (Table 3). Classes in the whole collection at a frequency lower than 1% were grouped together; such was the case for the America, Europe and Oceania continents of origin, and kafir intermediates in the race classification. Passport information and racial distribution in the core P and the landrace collection were not significantly different. As with the morpho-agronomic traits, the core P gave a faithful image of the landrace collection for geographical and racial distribution. Conversely, core C and core L differed significantly from the landrace collection for latitudinal, continental (except Asia for the core L) and racial distributions (except C, GB, GD for the core C and GB, GD and DC for the core L).

The three core collections were compared for the qualitative characters, quantitative characters, passport data and racial distribution (Table 4). From the previous comparison of core P and the landrace collection, no significant difference was found. Thus, core P is assumed to adequately represent the landrace collection in its comparisons with core C and L. As both core C and core L differed from the landrace collection, an assessment was made only on the differences between these two core collections. Core C and core L did not differ significantly for the qualitative characters. Core C and core L differed significantly for days to flowering in kharif, plant height in the two seasons, and one hundred seed weight (Table 4). Core C and core L were not statistically different in their distribution over the continents. However, significant differences were observed for the distribution of landraces in the tropical and temperate latitudinal classes (LR-2 and LR-3) and for the distribution of the race guinea.

## DISCUSSION

As suggested by Brown (1989b) in most cases, the stratification of the base collection has been the first step to constitute core collections (Brown, 1989b; Erskine and Muehlbauer, 1991; Spagnoletti Zeuli and Qualset, 1993; Basigalup et al., 1995; Diwan et al., 1995; Cordeiro et al., 1995; Crossa et al., 1995; van Hintum et al., 1995; Prasada Rao and Ramanatha Rao, 1995; Tohme et al., 1995; Balfourier et al., 1998; Iguarda et al., 1998; Ortiz et al., 1998; Huamán et al. 1999; Skinner et al., 1999). In sorghum, Prasada Rao and Ramanatha Rao (1995) stratified the base collection from a Principal Component Analysis based on seven morphological traits. In our study the base collection was stratified according to photoperiod sensitivity and such clustering was discussed in Grenier et al. (submitted). In the second step, either only one sampling strategy was applied (Prasada Rao and Ramanatha Rao, 1995), or comparisons between random sampling strategies were reported (Brown, 1989b; Erskine and Muehlbauer, 1991; Spagnoletti Zeuli and Qualset, 1993; Cordeiro et al., 1995; Diwan et al., 1995; Galwey, 1995; van Hintum et al., 1995). In the latter case, the main conclusions drawn from these studies were that a better representation of the pattern of



variation present in the total collection was obtained with the proportional sampling. The range of variation present in the whole collection was maximized when constant or logarithmic sampling were used. In the simulation studies of Schoen and Brown (1993), Schoen and Brown (1995), and Bataillon et al. (1996) and in the hypothetical populations of Yonezawa et al. (1995), the proportional sampling appeared as the optimal strategy, although differences between it and the logarithmic sampling were small.

Differences in representation between core collections could result from differences in the stratification of the base collection as it partitioned the total diversity. The stratification can lead to near equal or unequal clusters size. When group sizes are equal, the three random sampling procedures will likely be equivalent. For unequal group size, Brown (1989b) found in the *Glycine tomentella* core collection that when the rarest variants were classified in the smaller groups, then the constant strategy sampled the highest number of types or alleles per locus. Conversely, if the rarest variants occurred in the largest groups, the proportional sampling identified the best core. Brown (1989b) concluded the logarithmic strategy was a good compromise that both lowered the representation from the largest clusters and sampled the rare variants. In addition, for the Brazilian cassava collection, where areas of high diversity were poorly represented in the whole collection, the logarithmic strategy was recommended (Cordeiro et al., 1995). In a similar manner, the logarithmic strategy better sampled the under-represented countries in the durum wheat collection (Spagnoletti Zeuli and Qualset, 1993). The logarithmic strategy seems to reduce the weight given to the larger groups, which more likely have a high level of genetic redundancy. Also, it could be used to increase the probability of sampling rare alleles that confer wide or local adaptation.

In fact, the choice of the sampling strategy greatly depends on the objectives given to the core collection.

When the objective of defining a core collection and the properties of the three random sampling strategies are considered, two options are possible.

When the maintenance of a collection has become unmanageable because of its increasing size, the main objective of the core is to concentrate efforts on a smaller set that faithfully represents the diversity of the entire collection. In this situation, and when characterization data are non available, the P strategy would be the best to give a reliable image of the landrace collection as it was shown in our study.

When the core is to be used by the breeders for genetic enhancement, it should contain the maximum amount of the diversity present in the base collection and facilitate access to specific traits under-represented in the base collection. For this objective, the C or L strategy would be the best approach. In particular situations, L sampling is preferred to C sampling. For example, the logarithmic sampling is recommended to better sample the rarest variants from the smaller groups (Brown, 1989b on perennial wild relatives of soybean; Spagnoletti Zeuli and Qualset, 1993 on durum wheat; Cordeiro et al., 1995 on Brazilian cassava; Galwey, 1995 on beans).

In the present work, the constant and logarithmic core collections had different patterns of diversity when compared to the landrace collection. For sorghum breeders, photoperiod insensitive landraces are important for short-term genetic enhancement in the temperate latitude. Furthermore, photoperiod sensitivity is an important trait for tropical breeding programs. Indeed, it insures yield stability both for quantity and quality aspect particularly for cultivars grown under high climatic constraints as in Western Africa (Chanterreau et al., 1998; Kouressy et al., 1998; Trouche et al., 1998;). Thus, there is a need to keep the widest range of photoperiod sensitivity to retain a wide range of adaptation to specific eco-geographical zones in a core collection. The logarithmic core retained an



increased number of the photoperiod insensitive landraces and highly photoperiod sensitive landraces that would serve as an important source of broadly adapted genetic background for breeding programs.

#### ACKNOWLEDGMENTS

We acknowledge M. Deu, for her contributive remarks and review of the manuscript, and ICRISAT-GRD staff for providing data information.

#### REFERENCES

- Appa Rao, S., K.E. Prasada Rao, M.H. Mengesha, and V. Gopal Reddy. 1996. Morphological diversity in sorghum germplasm from India. *Gen. Res. Crop Evol.* 43:559-567.
- Balfourier, F., G. Charmet, J.-M. Prosperi, M. Goulard, and P. Monestiez. 1998. Comparison of different spatial strategies for sampling a core collection of natural populations of fodder crops. *Genet. Sci. Evol.* 30 (Suppl.1):215-235.
- Basigalup, D.H., D.K. Barnes, and R.E. Stuker. 1995. Development of a core collection for perennial *Medicago* plant introduction. *Crop Sci.* 35:1163-1168.
- Bataillon, T. 1994. Comparaison de diverses stratégies d'échantillonnage pour la constitution de core collections de ressources génétiques végétales (études par simulation informatique). Mémoire de DEA, Institut National Agronomique, Paris.
- Bataillon, T., J.L. David, and D.J. Schoen. 1996. Neutral genetic markers and conservation genetics: simulated germplasm collections. *Genetics* 144:400-417.
- Brown, A.H.D. 1989a. The case of core collections. p. 135-156. *In* A. H. D. Brown et al. (ed.) The use of plant genetic resources. Cambridge Univ. Press, Cambridge, UK.
- Brown, A.H.D. 1989b. Core collections: a practical approach to genetic resources management. *Genome* 31:818-824.
- Brown, A.H.D. 1995. The core collection at the crossroads. p. 3-19. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.
- Chantreau, J., C. Luce, M.A.G. Hamada, and G. Trouche. 1998. L'utilisation des sorghos guinea en sélection dans le programme conjoint ICRISAT-CIRAD sur le sorgho. p. 65-72. *In* A. Ratnadass et al. (ed.) Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre. Actes de l'atelier de restitution du programme conjoint sur le sorgho ICRISAT-CIRAD. Bamako, Mali. Collection colloques, Montpellier, CIRAD-CA.
- Cordeiro, C.M.T., E.A.V. Morales, P. Ferreira, D.M.S Rocha, I.R.S. Costa, A.C.C. Valois, and S. Silva. 1995. Towards a Brazilian core collection of cassava p. 155-168. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.
- Crossa, J., I.H. DeLacy, and S. Taba. 1995. The use of multivariate methods in developing a core collection. p. 77-92. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.
- Diwan, N., M.S. McIntosh, and G.R. Baughan. 1995. Methods of developing a core collection of annual *Medicago* species. *Theor. Appl. Genet.* 90:755-761.
- Erskine, W., and F.J. Muehlbauer. 1991. Allozyme and morphological variability, outcrossing rate and core collection formation in lentil germplasm. *Theor. Appl. Genet.* 83:119-125.
- Galwey, N.W. 1995. Verifying and validating the representativeness of a core collection. p. 187-198. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.
- van Hintum, Th.J.L., R. Von Bothmer, and D.L. Visser. 1995. Sampling strategies for composing a core collection of cultivated barley (*Hordeum vulgare* s. lat.) collected in China. *Hereditas* 122:7-17.
- Holbrook, C.C., W.F. Anderson, and R.N. Pittman. 1993. Selection of a core collection from the U.S. germplasm collection of peanut. *Crop. Sci.* 33:859-861.
- Huamán Z., C. Aguilar, and R. Ortiz. 1999. Selecting a Peruvian sweetpotato core collection on the basis of morphological, eco-geographical, and disease and pest reaction data. *Theor. Appl. Genet.* 98: 840-844.
- Hutcheson, K. 1970. A test for comparing diversities based on the Shannon formula. *J. Theor. Biol.* 29:151-154.
- Igartua, E., M.P. Gracia, J.M. Lasa, B. Medina, J.L. Molina-Cano, J.L. Montoya, and I. Romagosa. 1998. The Spanish barley core collection. *Gen. Res. Crop Evol.* 45:475-481.
- Kouressy, M., M. Ouattara, and M. Vaksman. 1998. Importance du photopériodisme chez les sorghos tropicaux, conséquences pour un programme de sélection. p. 49-54. *In* A. Ratnadass et al. (ed.)

- Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre. Actes de l'atelier de restitution du programme conjoint sur le sorgho ICRISAT-CIRAD. Bamako, Mali. Collection colloques, Montpellier, CIRAD-CA.
- Li, Y., S. Wu, Y.S. Cao, and X. Zhang. 1996. A phenotypic diversity analysis of foxtail millet (*Setaria italica* (L.) P.Beauv.) landraces of Chinese origin. *Gen. Res. Crop Evol.* 43:377-384.
- Ortiz, R., E.N. Ruiz-Tapia, and A. Mujica-Sanchez. 1998. Sampling strategy for a core collection of Peruvian quinoa germplasm. *Theor. Appl. Genet.* 96:475-483.
- Prasada Rao, K.E., and V. Ramanatha Rao. 1995. The use of characterisation data in developing a core collection of sorghum. p. 109-116. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.
- Schoen, D.J., and A.H.D. Brown. 1993. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proc. Natl. Acad. Sci. (USA)* 90:10623-10627.
- Schoen, D.J., and A.H.D. Brown. 1995. Maximising allelic diversity in core collections of wild crop relatives: the role of genetic markers. p. 55-76. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.
- Skinner D.Z., G.R. Bauchan, G. Auricht, and S. Hughes. 1999. A method for the efficient mangement and utilization of large germplasm collections. *Crop Sci.* 39: 1237-1242.
- Spagnoletti Zeuli, P. L., and C.O. Qualset. 1993. Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. *Theor. Appl. Genet.* 87:295-304.
- StatSoft. 1997. Statistica for Windows. USA, StatSoft, Tulsa, USA.
- Tohme, J., P. Jones, S. Beebe, and M. Iwanaga. 1995. The combined use of agroecological and characterisation data to establish the CIAT *Phaseolus vulgaris* core collection. p. 95-107. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.
- Trouche, G., M. Vaksman, J. Chantereau, M. Kouressy, H.D. Maiga, and C. Barro. 1998. Etude du déterminisme génétique du photopériodisme des sorghos guinea. p. 37-48. *In* A. Ratdnadass et al. (ed.) Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre. Actes de l'atelier de restitution du programme conjoint sur le sorgho ICRISAT-CIRAD, Bamako, Mali, Collection colloques, Montpellier, CIRAD-CA.
- Vaughan, D.A. 1991. Choosing rice germplasm for evaluation. *Euphytica* 54:147-154.
- Yonezawa, K., T. Nomura, and H. Morishima. 1995. Sampling strategies for use in stratified germplasm collections. p. 35-53. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.



Table 1. Descriptive statistics for the ten quantitative characters: Mean ( $\pm$  standard error), range values, and variance for the sorghum landrace collection and for the core collections. Levene's test of homogeneity of variances and Newman-Keuls' test (N-K test) for means comparison between each core and the sorghum landrace collection.

	Flowering date in dry season (days)	Flowering date in rainy season (days)	Basal tillers (No.)	Plant height in dry season (cm)	Plant height in rainy season (cm)	Peduncle exertion (cm)	Panicle length (cm)	Panicle width (cm)	Grain size (mm)	100-seed weight (g)
<b>Landrace collection</b>										
Mean $\pm$ se	72.64 $\pm$ 0.10	97.05 $\pm$ 0.19	2.08 $\pm$ 0.01	219.02 $\pm$ 0.37	342.50 $\pm$ 0.59	16.55 $\pm$ 0.06	22.17 $\pm$ 0.06	9.42 $\pm$ 0.03	3.03 $\pm$ 0.00	3.08 $\pm$ 0.01
Range	36-136	35-199	1-20	55-465	80-655	0-71	2.5-86	1-80	0.8-5.6	0.32-8.92
Variances	216.56	845.34	0.89	3045.29	7921.58	86.54	73.84	21.89	0.27	1.06
<b>Core P</b>										
Mean $\pm$ se	72.70 $\pm$ 0.31	97.08 $\pm$ 0.61	2.07 $\pm$ 0.02	220.37 $\pm$ 1.15	344.34 $\pm$ 1.85	16.71 $\pm$ 0.19	22.35 $\pm$ 0.19	9.43 $\pm$ 0.10	3.04 $\pm$ 0.01	3.09 $\pm$ 0.02
Range	36-132	46-165	1-10	70-465	85-655	0-60	4-86	1.3-41	1-5.5	0.71-7.51
Variances	216.11	832.44	0.86	2986.47	7723.48	83.65	78.01	21.66	0.27	1.03
Levene's test	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
N-K test	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
<b>Core C</b>										
Mean $\pm$ se	70.47 $\pm$ 0.29	83.09 $\pm$ 0.62	2.11 $\pm$ 0.02	207.80 $\pm$ 1.13	290.05 $\pm$ 2.09	16.17 $\pm$ 0.19	23.44 $\pm$ 0.16	9.52 $\pm$ 0.10	2.96 $\pm$ 0.01	2.87 $\pm$ 0.02
Range	36-132	40-164	1-13	65-465	90-610	0-60	5-57	1.3-41	0.8-5.5	0.32-6.84
Variances	193.52	854.64	0.82	2874.97	9812.20	77.10	58.04	21.54	0.25	0.77
Levene's test	**	NS	*	*	**	**	**	NS	**	**
N-K test	-	**	-	-	-	-	-	NS	-	-
<b>Core L</b>										
Mean $\pm$ se	71.11 $\pm$ 0.29	86.62 $\pm$ 0.63	2.07 $\pm$ 0.02	213.32 $\pm$ 1.12	303.23 $\pm$ 2.09	16.48 $\pm$ 0.19	23.52 $\pm$ 0.16	9.63 $\pm$ 0.10	2.98 $\pm$ 0.01	2.92 $\pm$ 0.02
Range	36-132	41-159	1-10	75-410	90-600	0-49	2.5-57	1-51	1-5	0.36-7.96
Variances	193.85	895.42	0.72	2837.14	9806.50	78.76	59.71	22.84	0.23	0.79
Levene's test	**	NS	**	*	**	**	**	NS	**	**
N-K test	-	**	-	-	-	-	-	*	-	-

NS = non-significant test at  $p > 5\%$  level; \*, \*\* significant at the  $p < 0.05$  and  $p < 0.01$  probability levels, respectively; - when N-K test can not be performed because of heterogeneous variances.



Table 2. Chi-square test ( $\chi^2$ ) comparing the frequency distribution observed in each core collections (Core P, C and L) with those expected in 10% of the sorghum landrace collection for the nine qualitative characters.

Qualitative Traits	No. of character states	$\chi^2$ Core P	$\chi^2$ Core C	$\chi^2$ Core L
Leaf pigmentation	2	0.61 NS	0.75 NS	0.03 NS
Midrib color	4	7.62 NS	79.13	71.82
Panicle compactness	10	7.70 NS	134.38	109.51
Glume color	13	8.01 NS	24.85	25.15
Glume covering	5	0.20 NS	174.11	151.13
Grain color	13	13.82 NS	145.26	101.00
Endosperm texture	5	2.93 NS	47.31	48.38
Grain lustre	2	0.66 NS	8.53	1.17 NS
Subcoat presence	2	2.16 NS	24.73	8.91

Leaf pigmentation; tan, pigmented - Midrib color; white, dull, yellow, brown - Panicle compactness; semi-compact oval, compact oval, semi compact elliptic, compact elliptic, semi loose drooping branches, loose drooping branches, very loose drooping branches, semi loose stiff branches, loose stiff branches, very loose stiff branches - Glume color; white, straw, yellow, gray, light brown, brown, partly brown, purple, partly purple, black, light red, red, red-brown - Glume covering; from uncovered (1) to fully covered (5) - Grain color; white, chalky white, white and red, gray, straw, yellow, light red, light brown, red, brown, red-brown, purple, black - Endosperm texture; from corneous (1) to floury (5) - Grain lustre; lustrous, non-lustrous - Subcoat; presence, absence.

NS : non significant difference at  $p < 0.05$  probability level.

Table 3. Chi-square test ( $\chi^2$ ) comparing the frequency distribution observed in the core collections with those expected in 10% of the sorghum landrace collection for the passport data (continents of origin and latitudinal range (LR)) and the race classification (the five basic race plus the intermediates).

	$\chi^2$ Core P	$\chi^2$ Core C	$\chi^2$ Core L
Africa	0.50 NS	5.25	6.14
Asia	0.41 NS	8.21	1.50 NS
America + Oceania + Europe	2.83 NS	538.43	337.26
LR-1 : 10°S-10°N	0.03 NS	144.60	116.83
LR-2 : 20°S-10°S, 10°N-20°N	0.19 NS	214.59	121.23
LR-3 : 30°S-20°S, 20°N-30°N	0.41 NS	1024.72	589.98
LR-4 : 40°S-30°S, 30°N-40°N	0.93 NS	1426.27	1093.69
B: bicolor	≡0.00 NS	131.48	74.66
C: caudatum	0.92 NS	2.43 NS	8.17
D: durra	0.09 NS	104.19	80.65
G: guinea	1.40 NS	75.54	37.88
K: kafir	1.92 NS	693.90	549.37
CB: caudatum-bicolor	1.60 NS	41.22	35.60
DB: durra-bicolor	0.30 NS	13.47	10.96
GB: guinea-bicolor	0.41 NS	2.91 NS	1.63 NS
DC: durra-caudatum	0.10 NS	13.44	3.84 NS
GC: guinea-caudatum	0.43 NS	21.62	22.30
GD: guinea-durra	0.35 NS	3.75 NS	0.74 NS
Intermediate kafir	1.53 NS	310.52	165.71

NS = non significant difference at  $p < 0.05$  probability level.

Table 4: Differences between the three random stratified sorghum core collections when compared for the ten quantitative characters, the nine qualitative characters, the passport data and the racial distribution. Different letters were used when the comparisons between the core collections were significantly different at  $p < 0.05$  probability level. Letter A refers to similar distribution as in the sorghum landrace collection, letter B and C refers to significantly different distributions to the landrace collection.

Traits	Core P	Core C	Core L
Days to flowering in rabi	A	B	B
Days to flowering in kharif	A	B	C
Basal tillers	A	B	B
Plant height in rabi	A	B	C
Plant height in kharif	A	B	C
Peduncle exertion	A	B	B
Panicle length	A	B	B
Panicle width	A	A, B	B
Grain size	A	B	B
100-seed weight	A	B	C
Leaves pigmentation	A	A	A
Midrib color	A	B	B
Panicle compactness	A	B	B
Glume color	A	B	B
Glume covering	A	B	B
Grain color	A	B	B
Endosperm texture	A	B	B
Grain lustre	A	B	A, B
Subcoat presence	A	B	B
Africa	A	B	B
Asia	A	B	A, B
America + Oceania + Europe	A	B	B
LR-1 : 10°S-10°N	A	B	B
LR-2 : 20°S-10°S, 10°N-20°N	A	B	C
LR-3 : 30°S-20°S, 20°N-30°N	A	B	C
LR-4 : 40°S-30°S, 30°N-40°N	A	B	B
B: bicolor	A	B	B
C: caudatum	A	A, B	B
D: durra	A	B	B
G: guinea	A	B	C
K: kafir	A	B	B
CB: caudatum-bicolor	A	B	B
DB: durra-bicolor	A	B	B
GB: guinea-bicolor	A	A	A
DC: durra-caudatum	A	B	A, B
GC: guinea-caudatum	A	B	B
GD: guinea-durra	A	A	A
Intermediate kafir	A	B	B



Assessment of Genetic Diversity in Three Subsets Constituted from the ICRISAT Sorghum Collection Using Random vs. Non-Random Sampling Procedures. A - Using morpho-agronomical and passport data.

Grenier C, Bramel-Cox PJ, Noirot M, Prasada Rao KE, Hamon P

*Accepté pour publication dans Theoretical and Applied Genetics*

---

ASSESSMENT OF GENETIC DIVERSITY IN THREE SUBSETS CONSTITUTED FROM THE ICRISAT  
SORGHUM COLLECTION USING RANDOM VS. NON-RANDOM SAMPLING PROCEDURES.

A - USING MORPHO-AGRONOMICAL AND PASSPORT DATA

C. Grenier, P.J. Bramel-Cox, M. Noirot, K.E. Prasada Rao, P. Hamon

Cécile Grenier, Centre de Coopération Internationale en Recherche Agronomique pour le  
Développement (CIRAD), BP 5035, 34032 Montpellier Cedex 1, France

Paula J. Bramel-Cox, International Crops Research Institute for the Semi-Arid Tropics  
(ICRISAT), Patancheru, 502 324 A.P., India

Michel Noirot, Institut de Recherche pour le Développement (IRD), BP 5045, 34032  
Montpellier Cedex 1, France

K.E. Prasada Rao, Academy of Biological Sciences, Visakhapatnam, 530 020 A.P., India

Perla Hamon, Université Montpellier III / IRD, route de Mende, 34199 Montpellier Cedex 5,  
France

Corresponding author: Cécile Grenier

ABSTRACT

A large collection, such as the sorghum (*Sorghum bicolor* [L.] Moench) landrace collection held at International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), represents a challenge for maintenance of both the accessions and the information documented on the germplasm collection. The accessibility and knowledge of the landrace collection are the essential factors for an efficient utilization of the genetic resources by both breeders and farmers. Different sampling strategies, either random or non-random were proposed to obtain subsets of reduced size (core collection). Three subsets were established; a random sampling within stratified collection (Logarithmic strategy: L); a sample based upon morpho-agronomic diversity (Principal Component Score strategy: PCS); and a sample based upon an empirical knowledge of sorghum (Taxonomic strategy: T). Comparisons of these three samples for morpho-agronomic characterization and passport information were assessed to determine the impacts on the phenotypic diversity. For their overall diversity, the three subsets did not differ as shown with the two-dimensional representation of the morpho-agronomic diversity and the Shannon-Weaver diversity indices. When comparison for morpho-agronomic and passport data were considered, the PCS subset looked similar to the entire landrace collection. The L subset showed differences for characters associated with the photoperiod reaction that was considered in the stratification of the collection. The T subset was the most distinct from the entire landrace collection as it over represented the landraces selected by farmers for specific uses and it covered the widest range of geographical adaptation and morpho-agronomic characteristics.

Key words: Core collection - Sorghum - Morpho-agronomic diversity - Sampling strategies

## INTRODUCTION

*Ex-situ* conservation of biodiversity can result in large collections that are difficult to characterize, evaluate, utilize and maintain. An important task for curators is to find a way to preserve the widest range of genetic diversity within crop species as well as to improve the knowledge and the utilization of the genetic resources. To alleviate management difficulties, the identification and use of core collections has been suggested. One of the objectives of this subset is to reduce the collection to a manageable size that will capture the widest range of the genetic diversity. Different sampling strategies have been proposed since Frankel first suggested the concept of core collection in 1984 (*in* Brown 1989a). Brown (1989b) proposed the use of random sampling strategies among a stratified collection. He assumed over 70% of the alleles present in the whole collection would be retained with a 10% of the total collection's size based upon the theory of selectively neutral alleles. For several plant species, core collections have been established and compared for their effectiveness to maintain the diversity of the base collection (Spagnoletti Zeuli and Qualset 1993; Bataillon 1994; Cordeiro et al. 1995; Diwan et al. 1995; Galwey 1995; van Hintum et al. 1995; Igartua 1998). For sorghum, Grenier et al. (b submitted) found that proportional sampling gave a better picture of the diversity of the base collection than constant or logarithmic sampling.

With an objective to retain the maximum genetic diversity from the whole base collection in a manageable working collection, several non-random sampling strategies have been suggested. The principal component score (PCS) strategy, which is based on phenotypic characters, has been applied to identify a subset that will maximize the representation of the phenotypic variability of the base collection (Mahajan 1996; Noirot et al. 1996; Balfourier et al. 1998; Bisht et al. 1998; Hamon et al. 1998). Alternatively, a subset could be identified based upon an empirical knowledge of the collection. This latter subset would be established with an understanding of the biology of the crop, its history, origins, the specific uses of its landraces, their relative importance, etc. Thus, both the phenotypic diversity as well as the crops' evolution is taken into account when this subset is selected.

The collection of cultivated sorghum maintained at ICRISAT contains more than 36,000 accessions. The objective of this study was to identify a subset or core collection that retained the widest range of the phenotypic diversity from this large collection. Three different sampling strategies, random and non-random, were used to establish the subsets. The three subsets were compared with each other and to the entire landrace collection based upon the morpho-agronomic and the passport data.

## MATERIAL AND METHODS

### SUBSETS CONSTITUTION

The entire sorghum landrace collection maintained at ICRISAT is completely described for morphological and passport data. This collection contains accessions acquired from a latitudinal range of 40°north to 40°south. The sorghum landrace collection of 22,473 accessions was used to establish two of the three subsets. The third subset was based upon the landrace collection plus some very recent additions to the gene bank.

Characterization data was recorded at Patancheru, India (17°25'N latitude and 78°E longitude) for 18 morpho-agronomic characters from 1975 to 1996 during the two typical seasons in India, i.e. the rabi season, and the kharif season (Appa Rao et al. 1996). Rabi is the dry season with short day lengths (September to April), while kharif is the rainy season with longer day duration (June to December). Ten quantitative characters were recorded. Days to 50% flowering were recorded as the mean emergence date to the date when 50% plants have started flowering. Plant height was observed at maturity and recorded as the length of the main stalk (cm). These two characters have been recorded in the kharif and rabi seasons, FLR, PHTR and FLK, PHTK, respectively. Peduncle exertion, PEDEX (cm), panicle length,



EHLG (cm), panicle width, EHWD (cm) and number of basal tillers, BT, were measured on plants in the field. After harvest, grain size GRS (mm) and one hundred seed weight SWT (g) were measured. The eight qualitative characters noted were midrib pigmentation, shape of the inflorescence, glume and grain color, glume covering, type of grain (corneous and lustrous), and presence or absence of subcoat. Passport information used included race classification into the five basic races (bicolor, caudatum, durra, guinea and kafir) and 10 intermediate forms (hybrid combination of the major races), according to Harlan and de Wet (1972), as well as the continent and latitudinal class of the country of origin.

The logarithmic subset (L) was sampled from a previously established logarithmic core collection that contained 2,247 accessions (Grenier et al. b submitted) which were previously distributed into four clusters for photoperiod sensitivity. A 10% random sampling (using the random generator in Statistica (StatSoft 1997)) was taken based on the logarithm of the number of accessions within each cluster. This sampling gave a subset of 225 accessions.

From the entire landrace collection, 2,000 randomly identified accessions were used to define the PCS subset from the principal component scoring strategy (Noirot et al. 1996) (PCS strategy 'corequant' obtained from <http://www.mpl.ird.fr/genetrop/rg/pcss/index.htm>). Thus, a principal component analysis was performed from the ten quantitative characters plus two ordinal characters i.e. glume covering and endosperm texture after standardization. The first 229 accessions with the highest relative contribution to the generalized sum of square were then selected. They represented 31% of the phenotypic diversity found in the 2,000 accessions.

The taxonomic subset (T) was selected based on empirical knowledge of the sorghum collection held by K.E. Prasada Rao. He was the senior botanist at ICRISAT who had been involved in the acquisition of the sorghum collection as well as in its characterization, evaluation, and conservation. He selected landraces that had a particular character of interest or originated from a specific region. Thus, a subset of 246 accessions was selected.

#### DATA ANALYSES

Principal component analysis is applied using the ten standardized quantitative characters from the landrace collection. The first and second principal component scores of the accessions identified in each subset are plotted in this reduced two-dimensional space.

The Shannon-Weaver Index is calculated from the eight qualitative characters plus the ten quantitative ones after their transformation into classes. Comparisons between the three subsets are made with t-test that used the approximated variance as defined in Grenier et al. (b submitted). For the ten quantitative characters, the entire landrace collection and the three subsets are compared through the Levene's test of homogeneity of variance and the Newman and Keuls's test for post-hoc mean comparisons.

Qualitative characters and passport data are compared between the three subsets and to the entire landrace collection with a chi-square test. When possible, classes with less than five individuals are grouped together. For the racial distribution, the durra-bicolor, guinea-bicolor and guinea-durra races are grouped together, as well as the kafir intermediate forms. On the contrary, when the grouping does not reach the minimum value (American plus European plus Oceanic landraces) the classes are not considered.

#### RESULTS

##### *Comparison of the subsets to the entire landrace collection*

Multivariate analysis of the ten quantitative characters gives a representation of the morpho-agronomic diversity on a reduced space for the entire landrace collection and the three subsets (Figure 1). The first axis accounts for 29.6% of the variance and the principal component scores are influenced by days to flowering and plant height in both seasons. The

second axis accounts for 19.2% of the variance and grain size and hundred seed weight influence the principal component scores. Accessions sampled in the three subsets adequately represent the space that defined the main diversity of the entire landrace collection. Also, accessions from the subsets are plotted in the space where only a small proportion of the entire landrace collection is represented. When the three subsets are compared, the morpho-agronomic diversity of each subset looks similar.

In order to estimate the global phenotypic diversity levels, Shannon-Weaver Index is calculated for the three subsets and the entire landrace collection. Statistically similar morpho-agronomic diversity levels are obtained ( $H' = 1.261 \pm 0.111$  for the entire landrace collection and  $1.253 \pm 0.109$ ,  $1.214 \pm 0.106$ , and  $1.242 \pm 0.113$ , for the PCS, the L and T subset, respectively). High levels of diversity are also found with the standardized Shannon indices ( $SDI = 0.853 \pm 0.053$ ,  $0.853 \pm 0.048$ ,  $0.823 \pm 0.052$  and  $0.826 \pm 0.051$ , for entire landrace collection, the subsets PCS, L and T, respectively).

The morpho-agronomic diversity based upon the mean and variance of the entire landrace collection and each of the three subsets is compared for the ten quantitative characters (Table 1). In the PCS subset, plant height in rabi and hundred seed weight have larger variance, and the grain size is on average smaller than the entire collection. The L subset has higher variance for plant height in kharif, and smaller variance for basal tillers, panicle length, grain size and hundred seed weight. The mean value for days to flowering in kharif is also earlier in the L subset. In the T subset, variance for days to flowering in kharif, panicle length, grain size and hundred seed weight is increased, while the variance for days to flowering and plant height in rabi is reduced. Also, the mean values for basal tillers and peduncle exertion are smaller in this T subset.

Morpho-agronomic diversity of the eight qualitative characters is also compared between the entire landrace collection and the three subsets (Table 2). For all characters, the PCS subset distribution does not differ significantly from the entire landrace collection. On the contrary, the T subset has a significantly different distribution for the all characters. The L strategy differs significantly for panicle compactness, glume covering, grain color, and endosperm texture.

The distribution of landraces among the continents is not significantly different between the three subsets and the entire landrace collection (Table 3). For the latitudinal distribution, the PCS subset and the entire landrace collection does not differ while in the L subset the representation in the four latitudinal classes was different. The T subset differs from the entire landrace collection for all the latitude classes but the tropical latitude (latitude class 2).

Considering the racial distribution, only two races, guinea race and CB forms, are not distributed differently between the entire landrace collection and the three subsets. The DB-GB-GD group was significantly reduced in the PCS subset while the representation of the GC form is increased. The L subset differs significantly from the entire landrace collection; the representation of bicolor and kafir races as well as the GC forms and the kafir intermediates is increased, while the representation of durra race is decreased. The representation of bicolor, durra and kafir races as those of kafir intermediates increase in the T subset but the representation of caudatum race and DC form decrease.

#### *Comparison between the subsets*

The three subsets are compared for the ten quantitative characters with the variance and mean comparison tests (Table 1). When compared to the L subset, the PCS subset has larger variance for plant height in rabi and seed weight, higher mean value for days to flowering and plant height in kharif, and smaller mean value for the panicle length. Comparison between the PCS and the T subset shows a greater variance for days to flowering and plant height in rabi in the PCS subset. The T subset has a greater variance for



days to flowering in kharif, panicle length and grain size. When compared to the L subset, the T subset has higher variance for panicle length, grain size and hundred seed weight; higher mean value for days to flowering both in rabi and kharif, and plant height in kharif and a smaller mean value for peduncle exertion.

Qualitative characters are also compared between each subset through chi-square test (Table 2). The distribution for three characters is significantly different between the PCS and the L subset (panicle compactness, glume covering and endosperm texture). When T and PCS subsets are compared, five characters have different distribution (midrib color, glume covering, grain color, grain lustre and subcoat presence). Comparison between the T and the L subset reveals a different distribution for all characters but endosperm texture. Only the glume covering has significantly different distribution for every subset comparison. Also, the L subset differs from the two other subsets for the panicle compactness, and the T subset differs significantly from the two other subsets for midrib color, grain color, grain lustre and subcoat presence.

Distribution of accessions within the continent of origin is similar between the three subsets, while latitudinal distribution differ between each subset for the equatorial latitude (latitude class 1). Moreover, the PCS subset differs significantly from the two other subsets for the highly temperate latitudes (latitude class 4), and the L subset differs significantly from the two other subsets for the temperate latitude (latitude class 3).

The racial distribution is not significantly different between the three subsets for guinea race, the DB-GB-GD group and the CB forms. However, when compared to the two other subsets, the representation of the kafir race is reduced in the PCS subset. Also the PCS subset has a smaller representation of bicolor race and kafir intermediates than the L subset, as well as a higher representation of caudatum race and DC form than in the T subset. The L subset has a smaller number of accessions classified as durra race than the PCS and T subsets. The T subset contrasts from the two other subsets, as its representation of the GC forms is not increased.

## DISCUSSION

Global diversity, as depicted through multivariate analysis of the ten continuous characters recorded on the entire landrace collection, demonstrates that the global morpho-agronomic diversity is well represented in the three subsets and the morpho-agronomic diversity is not influenced by the random or non-random sampling strategies. These results are supported by the high and similar values of SDI calculated from the entire landrace collection and the three subsets.

However, comparative analyses for each character between the three subsets and the entire landrace collection illustrate several differences. The PCS strategy results in a subset with no change in the variances or means for most of the quantitative characters from that of the entire landrace collection. This observation can be explained by the normal distribution of the characters in the entire landrace collection. However, the PCS sampling modifies the distribution patterns of the quantitative characters as the curves are flattened along the X-axis in comparison to the entire landrace collection. Nevertheless, it is noteworthy that this sampling strategy does not affect the distribution of the characters that are not considered in the statistical procedure. In our study, this result is verified irrespective of the type of character, i.e. qualitative morpho-agronomic character or passport data. Few reports on the use of the PCS strategy have been published. The application of this strategy to two perennial trees (rubber tree and coffea) and two annual crops (rice and sorghum) found that the distributions of the qualitative characters in the subsets were not modified to any significant degree while the quantitative characters demonstrated changes in the variances and means (Hamon et al. 1998). When the distribution was considered, for example in coffee, whatever



the shape of the curve in the base collection, the effect of the PCS sampling was characterized by a more or less flattened curve. This change in the pattern of distribution was likely due to less redundancy in this core. Thus, the PCS strategy permitted both the elimination of redundancies and the retention of the diversity of the entire landrace collection in a subset of manageable size (1% of the entire landrace collection in our work).

Second, when the logarithmic (L subset) or empirical (T subset) are compared, several or all characters (morpho-agronomic characters, either quantitative or qualitative, as well as passport data) have a modified distribution from the entire landrace collection. These discrepancies between the three subsets could be explained by two hypotheses. The first is that the differences are due to the sampling strategy itself and reflect a bias in the sampling. In this case, it would be possible to observe differences for any other set of data considered (resistance to biotic or abiotic stress, biochemical or molecular markers). The other hypothesis is that these differences are due to the method used for the constitution of these particular subsets. This second hypothesis looks the most likely especially when the L sampling strategy is considered. Indeed, the L subset is constituted from the landrace collection after it has been stratified into four clusters according to photoperiod sensitivity. The characters with the modified statistical parameters are those associated to the photoperiod sensitivity. For example, variance for plant height recorded during kharif season is increased while the mean value is decreased. Mean value for days to flowering recorded during kharif season is also reduced. Thus, the representation of early and short plant types was increased in the L subset. The sampling was also biased toward landraces with loose panicle and stiff branches that are the main characteristics for the varieties classified as either bicolor or kafir races. The proportions of the bicolor, kafir and the kafir intermediates as well as the proportion of the temperate material are higher in the L subset than in the entire landrace collection. These races and latitudinal class are mainly composed by photoperiod insensitive landraces and they represent a small portion among the entire landrace collection (Grenier et al., a submitted). Thus, the representation of photoperiod insensitive landraces is clearly increased in the L subset in relation to the entire landrace collection.

Differences between the subset and the entire landrace collection in the representation of the morpho-agronomic diversity, the geographical distribution and the race classification are even more marked in the T subset. This subset was defined from empirical knowledge based on many years of fieldwork. His criteria integrated a very wide set of characteristics, i.e. phenotypic characters, eco-environmental adaptation, geographic origin and race classification. As all these traits were considered together in the selective sampling, this subset was expected to differ from the entire landrace collection. This expectation is observed since, when compared to the entire landrace collection, most of the quantitative characters have heterogeneous variance and all the qualitative characters are differently distributed. In addition, racial and geographical distributions are greatly modified in this subset.

In conclusion, a random strategy (L subset) and a non-random strategy (PCS and T subsets) seem to retain a good representation of the overall diversity from entire landrace collection. However, dependent upon the sampling procedure, statistical parameters of some or all characters are modified. Two hypotheses, bias introduced during the sampling and bias due to the specific constitution of the subsets, are proposed to explain our results. A further assessment to compare the genetic diversity of the three subsets with the use of molecular markers, such as microsatellites, will be undertaken. Furthermore, it would be noteworthy to consider non-neutral characters for the sampling of core collections and neutral characters (as molecular markers) to estimate the effect of the sampling.

## ACKNOWLEDGEMENTS

The authors would like to thank Dr. Rodomiro Ortiz at the ICRISAT Genetic Resources Enhancement Program, Patancheru, India and Dr. Serge Hamon at the Institut de Recherche pour le Développement Montpellier, France for their valuable comments on the manuscript. The experiments comply with the current laws of the country in which the experiments were performed.

## REFERENCES

- Appa Rao S, Prasada Rao KE, Mengesha MH, Gopal Reddy V (1996) Morphological diversity in sorghum germplasm from India. *Genetic Resources and Crop Evolution* 43: 559-567
- Balfourier F, Charmet G, Prosperi J-M, Goulard M, Monestiez P (1998) Comparison of different spatial strategies for sampling a core collection of natural populations of fodder crops. *Genet Sci Evol* 30 (Suppl.1): S215-S235
- Bataillon T (1994) Comparaison de diverses stratégies d'échantillonnage pour la constitution de *core collections* de ressources génétiques végétales (études par simulation informatique), Mémoire de DEA, Institut National Agronomique, Paris
- Bisht IS, Mahajan RK, Patel DP (1998) The use of characterisation data to establish the Indian mungbean core collection and assessment of genetic diversity. *Gen Res Crop Evol* 45: 127-133
- Brown AHD (1989 a) The case for core collections. In: Brown AHD, Frankel OH, Marshall DR, Williams JT (eds) *The use of plant genetic resources*, Cambridge, pp135-156
- Brown AHD (1989 b) Core collections: a practical approach to genetic resources management. *Genome* 31: 818-824
- Cordeiro CMT, Morales EAV, Ferreira P, Rocha DMS, Costa IRS, Valois ACC, Silva S (1995) Towards a Brazilian core collection of cassava. In Hodgkin T, Brown ADH, van Hintum TJL, Morales EAV (eds) *Core collections of plant genetic resources*. John Wiley & Sons, Chichester, UK, pp 155-168.
- Diwan N, McIntosh MS, Bauchan GR (1995) Methods of developing a core collection of annual *Medicago* species. *Theor Appl Genet* 90: 755-761.
- Galwey NW (1995) Verifying and validating the representativeness of a core collection. In Hodgkin T, Brown ADH, van Hintum TJL, Morales EAV (eds) *Core collections of plant genetic resources*. John Wiley & Sons, Chichester, UK, pp 187-198.
- Grenier C, Bramel-Cox PJ, Hamon P (submitted a) Core collection of the genetic resources of sorghum: I. Stratification based on eco-geographical data. *Crop Sci*
- Grenier C, Hamon P, Bramel-Cox PJ (submitted b) Core collection of the genetic resources of sorghum: II. Comparison of three random sampling procedures. *Crop Sci*
- Hamon S, Dussert S, Deu M, Hamon P, Seguin M, Glaszmann J-C, Grivet L, Chanterau J, Chevallier M-H, Flori A, Lashermes P, Legnate H, Noirot M (1998) Effects of quantitative and qualitative principal component score strategies on the structure of coffee, rubber tree, rice, and sorghum core collections. *Genet Sel Evol* 30 (Suppl.1): S237-S258
- Harlan JR, de Wet MJM (1972) A simplified classification of cultivated sorghum. *Crop Sci* 12: 172-176
- Hintum Th JL van, Von Bothmer R, Visser DL (1995) Sampling strategies for composing a core collection of cultivated barley (*Hordeum vulgare* s. lat.) collected in China. *Hereditas* 122: 7-17. ICRISAT (1996) *The world sorghum and millet economies. Facts, trends and outlook*. ICRISAT & FAO
- Igartua E, Gracia MP, Lasa JM, Medina B, Molina-Cano JL, Montoya JL, Romagosa I (1998) The Spanish barley core collection. *Gen Res Crop Evol* 45: 475-481
- Mahajan RK, Bisht IS, Agrawal RC, Rana RS (1996) Studies on South Asian okra collection: methodology for establishing a representative core set using characterization data. *Gen Res Crop Evol* 43: 249-255
- Noirot M, Hamon S, Anthony F (1996) The principal component scoring: a new method of constituting a core collection using quantitative data. *Genetic resources and Crop Evolution* 43: 1-6
- Spagnoletti Zeuli PL, Qualset CO (1993) Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. *Theor Appl Genet* 87: 295-304.
- StatSoft (1997) *Statistica for Windows*. StatSoft, USA.



Figure 1. Morpho-agronomic diversity in the two dimensional space defined from the principal component analysis on the ten quantitative characters and for the 22,473 accessions. Plain line refers to the distribution of 90% of the entire landrace collection and dotted line refers to the more external distribution of the entire landrace collection. Only the coordinated of the accessions sampled in the three subsets (Principal Component Score strategy: PCS, Logarithmic strategy: L, and Taxonomic strategy: T) are plotted.



Table 1. Descriptive statistics for the ten quantitative characters; mean ( $\pm$  standard error), and variance for the entire landrace collection (ELC) and the three subsets (Principal Component Score strategy: PCS, Logarithmic strategy: L, and Taxonomic strategy: T). Levene's test of homogeneity of variances and Newman and Keuls' test for means comparison (A) between each subset and the entire landrace collection, and (B) between the subsets.

(A) Sets	Flowering date in rabi (days)	Flowering date in kharif (days)	Basal tillers (No.)	Plant height in rabi (cm)	Plant height in kharif (cm)	Peduncle exertion (cm)	Panicle length (cm)	Panicle width (cm)	Grain size (mm)	100 seeds weight (g)
ELC mean $\pm$ se	72.64 $\pm$ 0.10	97.05 $\pm$ 0.19	2.08 $\pm$ 0.01	219.02 $\pm$ 0.37	342.50 $\pm$ 0.59	16.55 $\pm$ 0.06	22.17 $\pm$ 0.06	9.42 $\pm$ 0.03	3.03 $\pm$ 0.00	3.08 $\pm$ 0.01
var.	216.56	845.34	0.89	3045.29	7921.58	86.54	73.84	21.89	0.27	1.06
PCS mean $\pm$ se	72.43 $\pm$ 1.01	93.45 $\pm$ 1.90	2.10 $\pm$ 0.06	220.33 $\pm$ 4.13	334.50 $\pm$ 5.99	15.41 $\pm$ 0.58	21.47 $\pm$ 0.54	9.21 $\pm$ 0.28	2.95 $\pm$ 0.03	3.18 $\pm$ 0.08
var.	234.04	824.24	0.73	3905.92	8216.08	77.51	66.20	18.07	0.24	1.38
Levene	NS	NS	NS	*	NS	NS	NS	NS	NS	*
N-K	NS	NS	NS	-	NS	NS	NS	NS	*	-
L mean $\pm$ se	71.24 $\pm$ 0.92	84.98 $\pm$ 1.98	2.08 $\pm$ 0.05	223.89 $\pm$ 3.52	306.31 $\pm$ 6.35	16.40 $\pm$ 0.61	24.47 $\pm$ 0.51	9.92 $\pm$ 0.34	2.96 $\pm$ 0.03	2.93 $\pm$ 0.06
var.	191.15	882.62	0.63	2786.48	9068.47	84.89	57.67	26.41	0.20	0.68
Levene	NS	NS	*	NS	*	NS	**	NS	**	**
N-K	NS	**	-	NS	-	NS	-	NS	-	-
T mean $\pm$ se	75.13 $\pm$ 0.80	100.62 $\pm$ 1.96	1.92 $\pm$ 0.05	221.30 $\pm$ 3.23	334.65 $\pm$ 6.17	14.19 $\pm$ 0.55	23.25 $\pm$ 0.60	9.79 $\pm$ 0.31	3.26 $\pm$ 0.04	3.33 $\pm$ 0.08
var.	156.39	948.29	0.70	2566.46	9372.84	73.97	88.17	23.55	0.37	1.60
Levene	*	*	NS	*	NS	NS	*	NS	**	**
N-K	-	-	**	-	NS	**	-	NS	-	-

(B) Sets	Flowering date in rabi (days)	Flowering date in kharif (days)	Basal tillers (No.)	Plant height in rabi (cm)	Plant height in kharif (cm)	Peduncle exertion (cm)	Panicle length (cm)	Panicle width (cm)	Grain size (mm)	100 seeds weight (g)
PCS vs. L										
Levene	NS	NS	NS	*	NS	NS	NS	NS	NS	**
N-K	NS	**	NS	-	**	NS	**	NS	NS	-
PCS vs. T										
Levene	*	*	NS	**	NS	NS	*	NS	**	NS
N-K	-	-	*	-	NS	NS	-	NS	-	NS
L vs. T										
Levene	NS	NS	NS	NS	NS	NS	**	NS	**	**
N-K	**	**	NS	NS	**	*	-	NS	-	-

NS : non-significant test at  $p > 5\%$  level; \*, \*\* significant at the  $p < 0.05$  and  $p < 0.01$  levels, respectively.

Table 2. Comparison for the frequency distribution for the qualitative characters through  $\chi^2$  test (A) between the three subsets (Principal Component Score strategy: PCS, Logarithmic strategy: L, and Taxonomic strategy: T) and the entire landrace collection (ELC), and (B) between the three subsets.

Qualitative characters	No. of modalities	(A) Comparison to the entire landrace collection			(B) Comparison between subsets		
		PCS vs. ELC	L vs. ELC	T vs. ELC	PCS vs. L	PCS vs. T	L vs. T
Midrib color	2	3.22 NS	0.34 NS	28.06**	3.04 NS	20.82**	8.10**
Panicle compactness	4	0.62 NS	15.39**	16.73**	7.88*	7.76 NS	27.22**
Glume color	9	5.43 NS	15.07 NS	24.96**	11.01 NS	14.76 NS	23.64**
Glume covering	5	1.92 NS	31.68**	31.66**	11.68*	14.67**	13.86**
Grain color	7	4.46 NS	16.96**	55.91**	9.13 NS	36.86**	47.80**
Endosperm texture	5	1.59 NS	13.21*	15.32**	11.16*	9.43 NS	7.37 NS
Grain lustre	2	0.00 NS	0.94 NS	30.59**	0.52 NS	16.83**	11.31**
Subcoat presence	2	0.00 NS	0.13 NS	28.33**	0.06 NS	16.26**	18.07**

NS : non-significant test at  $p > 5\%$  level; \*, \*\* significant at the  $p < 0.05$  and  $p < 0.01$  levels, respectively.



Table 3: Frequency for the passport data and comparison between the entire landrace collection (ELC) and the subsets (PCS, L and T), and between the three subsets. The frequency distributions were compared through a chi-square test. Same letters showed not significant differences at  $p<0.05$  level. Number of accessions is given in brackets.

	ELC	PCS	L	T
Africa	(15652) a	(162) a	(151) a	(171) a
Asia	(6329) a	(64) a	(56) a	(66) a
Latitude 1: 10°N-10°S	(8919) a	(102) a	(47) b	(75) c
Latitude 2: 20°N-20°S	(10583) a	(96) a	(69) b	(102) a, b
Latitude 3: 30°N-30°S	(1843) a	(23) a, c	(64) b	(37) c
Latitude 4: 40°N-40°S	(1128) a	(8) a	(45) b	(32) b
bicolor	(718) a	(6) a, c	(16) b	(14) b, c
caudatum	(4893) a	(54) a	(37) a, b	(34) b
durra	(4545) a	(42) a, b	(22) c	(66) b
guinea	(3907) a	(33) a	(27) a	(43) a
kafir	(520) a	(7) a	(23) b	(21) b
CB	(760) a	(6) a	(13) a	(10) a
DB-GB-GD	(1620) a	(7) b	(12) a, b	(13) a, b
DC	(2713) a	(35) a	(23) a, b	(19) b
GC	(2399) a	(35) b	(39) b	(16) a
intermediate kafir	(398) a	(4) a, c	(13) b	(10) b, c

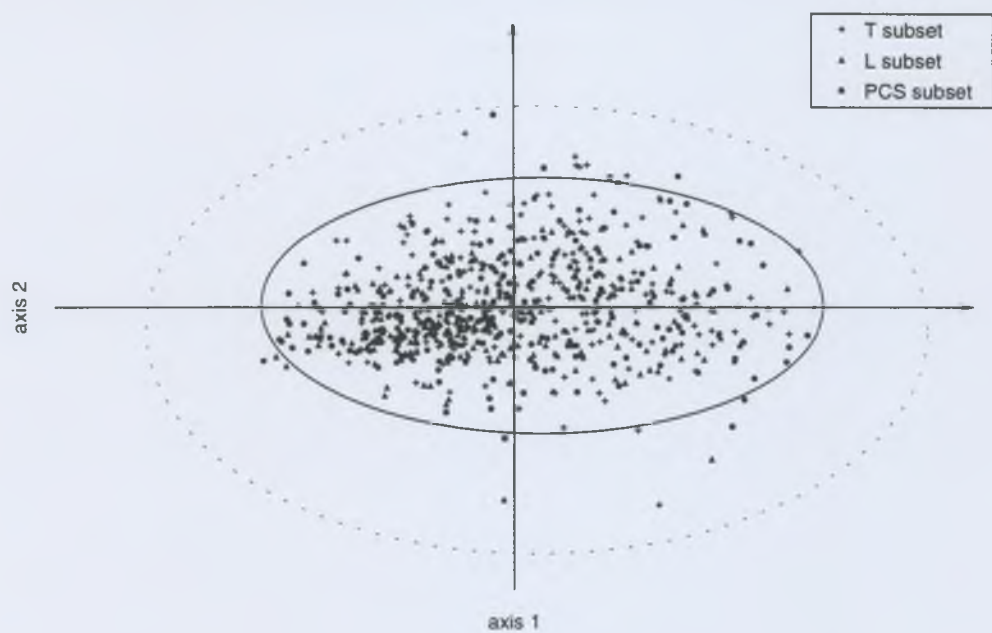


Figure 1

Assessment of Genetic Diversity in Three Subsets Constituted from the ICRISAT Sorghum Collection Using Random vs. Non-Random Sampling Procedures. B - Using molecular markers.

Grenier C, Deu M, Kresovich S, Bramel-Cox PJ, Hamon P

*Accepté pour publication dans Theoretical and Applied Genetics*

---



ASSESSMENT OF GENETIC DIVERSITY IN THREE SUBSETS CONSTITUTED FROM THE ICRISAT  
SORGHUM COLLECTION USING RANDOM VS. NON-RANDOM SAMPLING PROCEDURES.

B - USING MOLECULAR MARKERS

C. Grenier, M. Deu, S. Kresovich, P.J. Bramel-Cox, P. Hamon

Cécile Grenier and Monique Deu, Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), BP 5035, 34032 Montpellier Cedex 1, France; Paula J. Bramel-Cox, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, 502 324 A.P., India; Stephen Kresovich, Institute for Genomic Diversity, Cornell Univ., Ithaca, NY 14853-2703; Perla Hamon, Université Montpellier III/IRD, route de Mende, 34199 Montpellier Cedex 5, France

Corresponding author: Cécile Grenier

ABSTRACT

The large size of the sorghum (*Sorghum bicolor* (L.) Moench) landrace collection maintained by ICRISAT lead to the establishment of a core collection. Thus, three subsets of around 200 accessions were established from i) a random sampling after stratification of the entire landrace collection (L), ii) a selective sampling based on quantitative characters (PCS) and iii) a selection based on the geographical origin of landraces and the traits under farmers' selection (T). The assessment was done of the genetic diversity retained by each sampling strategy using the polymorphism at 15 microsatellite loci. The landraces of each subset were genotyped with three multiplex polymerase chain reactions (PCRs) of five fluorescent primer-pairs each with semi-automated allele sizing. The average allelic richness for each subset was equivalent (16.1, 16.3 and 15.4 alleles per locus for the subset PCS, L, and T, respectively). The average genetic diversity was also comparable for the three subsets (0.81, 0.77 and 0.80 for the subset PCS, L, and T, respectively). Allelic frequency distribution for each subset was compared with a chi-square test but few significant differences were observed. A high percentage of rare alleles (71 to 76% of 206 total rare alleles) was maintained in the three subsets. The global molecular diversity retained in each subset was not affected by a sampling procedure based upon phenotypic characters.

Key words: Core collection - Sorghum - SSRs - Genetic diversity

## INTRODUCTION

To assist in the management of a large collection, core collections have been defined to target maintenance and evaluation on a subset that retains a large part of the diversity encompassed in the entire collection (Brown 1989). Most core collections have been established from random procedures applied either directly on the initial set or on a stratified initial set. In this latter case, the stratification was often based upon biological or eco-geographical considerations. The degree of diversity in the core collection were essentially based on phenotypic values (Basigalup et al. 1995; Casler 1995; Charmet and Balfourier 1995; Cordeiro et al 1995; Diwan et al. 1995; Hodgkin 1995; Rincon et al. 1997; Balfourier et al. 1998; Ortiz et al. 1998; Huaman et al. 1999). Recently, some of these core collections went through specific evaluation procedures. For example, the core collection of US perennial alfalfa established by Basigalup et al. (1995) was assessed for acid soil tolerance (Bouton 1996). The results demonstrated that i) the diversity of the current core collection was sufficient to initially identify acid soil tolerant germplasm, and ii) that germplasm with exceptional acid soil tolerance will be difficult to isolate from the alfalfa collection. The same alfalfa core collection was assessed for forage quality variation and it provided a wider range of variation for forage quality traits than present in contemporary cultivars (Jung et al. 1997). A core subset established from a geographically stratified Peruvian quinoa collection and with a proportional sampling was screened for salt tolerance by Ruiz-Tapia et al. (1997) in Ortiz et al. (1999). The most tolerant accessions were found within a particular geographic origin and showed distinct morphological characteristics. From this evaluation for salt tolerance, the authors found that the core collection could serve as a starting point for further exploitation of the genetic resources available on the Peruvian quinoa genebank.

On sorghum (*Sorghum bicolor* (L.) Moench), core collections of 10% size have been established with random procedures and assessed for their phenotypic diversity (Grenier et al. submitted). Then, efforts were concentrated on three subsets of 1% size in order to assess the genetic diversity retained when random vs non-random sampling strategies are applied. Comparisons between the entire landrace collection and the three subsets (Grenier et al. submitted) found that the three subsets represented the overall phenotypic diversity of the entire landrace collection. Furthermore, they captured a high level of phenotypic diversity as measured by Shannon-Weaver Diversity indices. However, only one selective sampling did not affect the morpho-agronomic characters that were not previously considered in the sampling. Two hypotheses were formulated in order to explain such results. It was suggested that molecular markers could help in resolving such matter. Thus, the aim of this study was 1) to compare the genetic diversity of the three subsets using microsatellites 2) to assess the impact of the sampling strategy at the DNA diversity level and 3) to determine the rate of rare alleles retained in each procedure.

## MATERIAL AND METHODS

The logarithmic subset (L) is sampled from ICRISAT sorghum landrace collection that had been stratified into four clusters of photoperiod sensitivity. This sampling gives a subset of 225 accessions. The PCS subset is obtained using the principal component scoring strategy (Noirot et al. 1996) applied on a set of 2,000 randomly identified accessions. This subset includes 229 accessions. The taxonomic subset (T) has 246 accessions selected on their particular characters of interest and their specific geographic origin (Grenier et al. submitted).

Genomic DNA of each accession was extracted from on average 30 four days-old seedlings using the DNA extraction protocol described in Hoisington (1994). PCR amplification was carried out at USDA-ARS Plant Genetic Resources Conservation Unit's laboratory, Griffin, Ga, U.S.A.. Multiplex PCR of 15 SSRs loci (Table 1) was performed with fluorescently labeled primer pairs, as described in Dean et al. (1999). To check for



repeatability between gels, each gel run included the amplification products of a single plant DNA template from the breeding line BTx623. Automated sizing and labeled alleles were determined relative to an internal lane standard using ABI Genescan 672 software version 2.1. Limits of detection for fragments were arbitrarily chosen so that one repeat longer band with peak height lesser than 10% of the highest peak was not called. Thus, it eliminated bands that were due to the slippage of the *Taq* polymerase. As well, bands with an extra one base that resulted from the terminal deoxynucleotidyl transferase activity of *Taq* polymerase were not considered.

The genetic diversity index per locus was calculated as  $1 - \sum p_i^2$  where  $p_i$  is the frequency of the  $i^{\text{th}}$  SSR allele (Nei, 1972). The effective number of alleles per locus was obtained from the number of alleles whose frequency was higher than 5%. On these frequent alleles, chi-square tests were performed between the three subsets and to the global set.

## RESULTS

The rate of homogeneity on the global set (i.e. the three subsets all together) is variable between the 15 loci (Table 1). For example, 57% of the accessions of the global set are homogeneous at the locus Sb6-42, while 98% of the accessions are homogeneous at the locus Sb5-256. The rates observed are very similar between the three subsets; loci that revealed the highest or the less heterogeneity are the same in the three subsets.

For sorghum, heterogeneity within an accession could reflect heterozygosity at a locus and/or seed mixture. In this study it would be difficult to distinguish between these cases, thus hereafter we will only consider homogeneous accessions. With this condition, a total number of 289 alleles is recorded for the 15 loci (Table 2). In the global set, the mean allelic richness is 19.3 alleles per locus, and ranges from seven alleles at the locus Sb5-256 to 33 alleles at the locus Sb4-32. The gene diversity ranges between 0.71 and 0.93 excepted for the locus Sb5-256 that has a gene diversity value of 0.27.

Comparisons between the three subsets reveal similar tendencies. The mean allelic richness is of the same magnitude (16.3, 16.1 and 15.4 alleles per locus for the L, the PCS and the T subset, respectively). The genetic diversity is high in all three subsets. It ranges for 14 of the 15 loci from 0.72 to 0.93 (mean of 0.81) in the PCS subset, 0.66 to 0.92 (mean of 0.77) in the L subset and 0.74 to 0.93 (mean of 0.80) in the T subset. For one locus, Sb5-256, low and comparable genetic diversity values are found (0.23, 0.28 and 0.3 for the L, the T and the PCS subset, respectively).

When only the frequent alleles are considered, mean allelic richness for the global set is 5.5 alleles per locus, with a range of 2 (locus Sb5-256) to 9 alleles (locus Sb6-42) (Table 3). Once again, for each subset, the mean allelic richness is of same magnitude (5.7, 5.4 and 5.3 for PCS, L and T subset respectively). The chi-square tests between the three subsets and to the global set are performed only on the frequent alleles. When compared to the global set, the T subset does not have significantly different allelic distribution at any of the 15 loci. The PCS subset differs only at one locus (Sb1-10). The L subset has three loci (Sb4-15, Sb4-32 and Sb1-10) with a significantly different allelic distribution at  $p < 0.05$ ; one locus (Sb6-84) is in limit of significance at 5%. Comparisons between subsets show significant differences between the PCS and L subset for allelic distribution at 5 loci (Sb4-15, Sb4-32, Sb5-236, Sb1-10 and Sb6-84). Also, the PCS and T subset have 3 loci (Sb1-1, Sb6-42 and Sb4-72) with significantly different allelic distributions. The allelic distribution in the L and T subset is significantly different for four loci (Sb4-15, Sb4-32, Sb6-84, and Sb6-57).

In the global set, 206 alleles have a frequency lower than 5%. From these 206 rare alleles, 157 are present in the PCS subset from which 16 are found specifically in this subset. An equivalent number of rare alleles (158) is kept in the L subset and 13 of them are specific. In the T subset, 146 of the rare alleles are maintained, and 17 of them are specifically present



in this subset. Some of the rare alleles are also specifically lost in the subset: 31 in the taxonomic subset, 19 and 15 in the PCS and logarithmic subsets, respectively.

## DISCUSSION

Efficiency of *ex-situ* conservation and its use could be improved with greater genetic information on the germplasm. The study of 15 SSRs markers on three subsets of the sorghum landrace collection held at ICRISAT found that the homogeneity rate varied between 46 to 100% depending on the subset and the locus considered. The origin of heterogeneity in landraces of self-pollinated species needs to be considered for their secure management. Although sorghum is a preferentially autogamous crop, the out-crossing rate can range between 5 to 30% (Ollitrault 1987; Pedersen et al. 1998). As a result, some landraces could still represent traditional populations, although the possibility of seed mixtures either at the harvest or in the seed stocks cannot be excluded. Accession heterogeneity is a key point and required to be considered in a meaningful way. For genebank management, much care is necessary during the regeneration and evaluation process.

Molecular markers have an important role to play in many aspects of conservation (Karp et al., 1997). For instance they can be used for assessing genetic diversity in large collection. However, their uses need to be considered with both their cost and the status of the collection. For well-characterized collection, with a good extensive and valuable data set, morpho-agronomic characters can be considered as an efficient tool to assess genetic diversity. In this case, molecular markers can be great tools for fine resolution questions on plant genetic resources conservation (Kresovich et al. 1993). However, when collections are not fully characterized, or if genetic information provided by the characters are confounded by association or bias, then the molecular markers can be helpful for genetic diversity assessment. In any ways, the molecular studies should be considered as complementary to the phenotypic characterization.

In comparison with other markers, SSRs markers can be very appropriate tools to assess the genetic diversity in sorghum. Between two to four alleles per polymorphic locus were found with isozymes, RFLP, or RAPD markers (Ollitrault et al. 1989; Morden et al. 1989; Deu et al. 1994; de Oliveira et al. 1996; Deu et al. 1999). The 15 SSRs markers developed by Brown et al. (1996) revealed from two to six fragment size per polymorphic locus and a diversity index from 0.21 to 0.73 on 17 temperately and tropically adapted lines of sorghum (Brown et al. 1996). Moreover, the 15 microsatellite loci are widely spread on the sorghum genome and 14 of them have been mapped for nine of the ten sorghum linkage groups (Dean et al. 1999). With the same SSRs set, Dean and al. (1999) assessed diversity among 95 Orange accessions and found from three to 11 alleles per locus and a genetic diversity ranking 0.16 to 0.77 (Dean et al. 1999). In our study, using this SSRs set, from seven to 33 alleles per locus and a genetic diversity range of 0.71 to 0.93 (one locus at 0.27) were found. These differences between molecular markers in number of alleles per locus are in concordance with what is known on the mechanisms that generate variability for these different categories of markers e.g. punctual mutation, insertion/deletion for all the markers, plus polymerase slippage and/or unequal crossing-over for microsatellite markers.

The high values of allelic richness and genetic diversity found in the global set quantify the degree of genetic diversity in the entire landrace collection. These results suggest that the high phenotypic diversity of the sorghum landrace collection is associated with a high level of diversity at the DNA level. It is also possible to use such markers for further genotyping and elimination of redundancies as reported by Hokanson et al. (1998) on apple, by Chavarriaga-Aguirre et al. (1999) on cassava, and Dean et al. (1999) on the Orange sorghum accessions. The high values for genetic diversity here support the hypothesis that in

sorghum none these alleles have a selective advantage. For frequent alleles, the allelic frequencies are well distributed in the global set.

Comparisons between the three subsets highlighted similar genetic diversity retained by each sampling. From the previous comparison of the three subsets with morpho-agronomic data, we found same global phenotypic diversity, as measured with the Shannon diversity index. However, two hypotheses were considered to explain the discrepancy in morpho-agronomic distribution between the three sampling strategy i.e., bias introduced during the sampling and bias due to the specific constitution of the subsets (Grenier et al. submitted). Our present results of equal global molecular genetic diversity between the three subsets indicate similar efficiency of the three strategies to retain genetic diversity. However, allelic frequencies are found significantly different for some loci between the three subsets. Interestingly, the L subset differs from the PCS and T ones, for the same three loci. These results would support the hypothesis that the phenotypic characters that are influenced by the sampling are not affected because of a bias during the sampling. Indeed, in such a case, it would have been expected different allelic distribution for the three subsets at any of the locus. On the contrary, our results suggest that phenotypic characters as molecular markers could be affected as a result of their associations with the characters used for the constitution of the specific core collections. For example, if the stratification of the entire landrace collection was based on racial or geographic considerations, it would have been expected that the sampling does not affect the same characters than in Grenier's et al. (submitted) study, nor the same loci than in this study. Thus, could molecular markers be used to constitute a core collection that would retain the best sample of the agro-morphological diversity? This case could be considered for collections that were not yet evaluated and for which the cost of field analyses is much higher than these of molecular analyses. In the case of large collection such as the ICRISAT sorghum collection, it is not feasible to assess molecular makers on the entire collection because of the cost. To assess the potential in a limited manner, we considered the global set evaluated in this study as the target set and the five microsatellite loci for which data were the most complete as the sampling criteria (Sb4-121, Sb1-1, Sb6-84, Sb6-34 and Sb4-72). Then a selective sampling strategy on an initial set of 521 accessions with maximization of the molecular diversity (PCS strategy 'corequal' obtained from <http://www.mpl.ird.fr/genetrop/rg/pcss/index.htm>) was applied in order to establish a subset that represents 10% size of the target set. This subset includes 52% of total variability (data not shown). Comparisons of the phenotypic diversity between the subset and the target set based on the same characters than those used by Grenier et al. (submitted) found that the molecular PCS sampling does not induce changes in the distribution of the morpho-agronomic characters except for flowering days in rainy season. In these conditions, it would be interesting to consider an initial set that well represents the entire landrace collection as well as others sets of microsatellite loci. Furthermore, any association should be assessed with an investigation of the relationship between molecular markers, the specific morpho-agronomic characteristics and passport data.

## REFERENCES

- Balfourier F, Charmet G, Prosperi J-M, Goulard M, Monestiez P (1998) Comparison of different spatial strategies for sampling a core collection of natural populations of fodder crops. *Genet Sci Evol* 30 (Suppl.1): S215-S235
- Basigalup DH, Barnes DK, Stuker RE (1995) Development of a core collection for perennial *Medicago* plant introduction. *Crop Sci* 35: 1163-1168
- Bouton JH (1996) Screening the Alfalfa core collection for acid soil tolerance. *Crop Sci* 36: 198-200
- Brown AHD (1989 a) The case for core collections. In: Brown AHD, Frankel OH, Marshall DR, Williams JT (eds) *The use of plant genetic resources*, Cambridge, pp135-156
- Brown AHD (1989 b) Core collections: a practical approach to genetic resources management. *Genome* 31: 818-824



- Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzales-Candelas F, Kresovich S (1996) Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. *Theor Appl Genet* 93: 190-198
- Casler MD (1995) Patterns of variation in a collection of perennial ryegrass accessions. *Crop Sci* 35: 1169-1177
- Charmet G, Balfourier F (1995) The use of geostatistics for sampling a core collection of perennial ryegrass populations. *Gen Res Crop Evol* 42: 303-309
- Cordeiro CMT, Morales EAV, Ferreira P, Rocha DMS, Costa IRS, Valois ACC, Silva S (1995) Towards a Brazilian core collection of cassava. In: Hodgkin T, Brown AHD, Hintum ThJL van, Morales EAV (eds) *Core collections of plant genetic resources*. John Wiley & Sons, Chichester, UK, pp155-168
- de Oliveira A, Richter T, Bennetzen JL (1996) Regional and racial specificities in sorghum germplasm assessed with DNA markers. *Genome* 39: 579-587
- Dean RE, Dahlberg JA, Hopkins MS, Mitchell SE, Kresovich S (1999) Genetic redundancy and diversity among 'Orange' accessions in the U.S. national sorghum collection as assessed with simple sequence repeat (SSR) markers. *Crop Sci* 39: 1215-1221
- Deu M, Hamon P, Bonnot F, Chantereau J (1999) Le Sorgho. In: Hamon P (eds) *Diversité génétique des plantes tropicales cultivées*, CIRAD, Montpellier, France, pp 351-382
- Diwan N, Bauchan GR, McIntosh MS (1994) A core collection for the United States annual *Medicago* germplasm collection. *Crop Sci* 34: 279-285
- Dje Y, Forcioli D, Ater M, Lefebvre C, Vekemans X (1999) Assessing population genetic structure of sorghum landraces from North-Western Morocco using allozyme and microsatellite markers. *Theor Appl Genet* 99: 157-163
- Grenier C, Bramel-Cox PJ, Noirot M, Prasada Rao KE, Hamon P (submitted) Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs. non-random sampling procedures. A - Using morpho-agronomical and passport data. *Theor Appl Genet*
- Hamon S, Dussert S, Deu M, Hamon P, Seguin M, Glaszmann J-C, Grivet L, Chanterau J, Chevallier M-H, Flori A, Lashermes P, Legnate H, Noirot M (1998) Effects of quantitative and qualitative principal component score strategies on the structure of coffee, rubber tree, rice, and sorghum core collections. *Genet Sel Evol* 30 (Suppl.1): 237-258
- Hoisington D, Khairaliah M, Gonzales-de-Leon D (1994) *Laboratory protocols: CIMMYT Applied Molecular Genetics Laboratory*. Second edition
- Hokanson SC, Szewc-McFadden AK, Lamboy WF, McFerson JR (1998) Microsatellites (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus x domestica* borkh. core subset collection. *Theor Appl Genet* 97: 671-683
- Huaman Z, Aguilar C, Ortiz R. (1999) Selecting a Peruvian sweetpotato core collection on the basis of morphological, eco-geographical, and disease and pest reaction data. *Theor Appl Genet* 98: 840-844
- Igartua E, Gracia MP, Lasa JM, Medina B, Molina-Cano JL, Montoya JL, Romagosa I (1998) The Spanish barley core collection. *Gen Res Crop Evol* 45: 475-481
- Jung HG, Sheaffer CC, Barnes DK, Halgerson JL (1997) Forage quality variation in the U.S. alfalfa core collection. *Crop Sci* 37: 1361-1366
- Karp A, Kresovich S, Bhat KV, Ayad WG, Hodgkin T (1997) Molecular tools in plant genetic resources conservation: a guide to the technologies. IPGRI Technical Buletins No.2
- Kresovich S, Lamboy WF, Szewc-McFadden AF, McFerson J, Forsline PL (1993) Molecular diagnostics and plant genetic resources conservation. *Ag Biotech News and Information* vol.5 No.7 : 225N-258N
- Morden CW, Doebley, J. F., Schertz, K. F. (1989) Allozyme variation in old world races of *Sorghum bicolor* (Poaceae). *Amer J Bot* 76(2): 247-255
- Noirot M, Hamon S, Anthony F (1996) The principal component scoring: a new method of constituting a core collection using quantitative data. *Gen Res Crop Evol* 43: 1-6
- Ollitrault P (1987) *Evaluation génétique des sorghos cultivés (Sorghum bicolor L. Moench) par l'analyse conjointe des diversités enzymatique et morphophysiologique; relation avec les sorghos sauvages*. Doctorat es sciences, Paris-XII, Orsay
- Ollitrault P, Noyer J-L (1989) Polymorphisme enzymatique des sorghos. III - Identification et classification de variétés améliorées de l'IRAT. *L'agronomie tropicale* 45(1): 59-66
- Ortiz R, Ruiz-Tapia, EN, Mujica-Sanchez A (1998) Sampling strategy for a core collection of Peruvian quinoa germplasm. *Theor Appl Genet* 96: 475-483
- Ortiz R, Madsen S, Ruiz-Tapia EN, Jacobsen S-E, Mujica-Sanchez A, Christiansen JL, Stolen O (1999) Validating a core collection of Peruvian quinoa germplasm. *Gen Res Crop Evol* 46: 285-290
- Pedersen JF, Toy JJ, Johnson B (1998) Natural outcrossing of Sorghum and Sundangrass in the central great plains. *Crop Sci* 38: 937-939
- Pederson GA, Fairbrother TE, Greene SL (1996) Cyanogenesis and climatic relationships in U.S. white clover germplasm collection and core subset. *Crop Sci* 36: 427-433



- Rincon F, Johnson B, Crossa J, Taba S (1997) Identifying subsets of maize accessions by three-mode principal component analysis. *Crop Sci* 37: 1936-1943
- StatSoft. (1997) Statistica for Windows. USA, StatSoft, USA.
- Tohme J, Jones P, Beebe S, Iwanaga M (1995) The combined use of agroecological and characterisation data to establish the CIAT *Phaseolus vulgaris* core collection. In: Hodgkin T, Brown AHD, Hintum ThJL van, Morales EAV (eds) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK, pp95-107
- Tohme J, Gonzalez O D, Beebe S, Duque MC (1996) AFLP analysis of gene pools of a wild bean core collection. *Crop Sci* 36: 1375-1384
- Weir BS (1989) Sampling properties of gene diversity. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding, and genetic resources. Sinauer Associates Inc. Sunderland, Massachusetts, pp23-42

Table 1: Multiplex primer sets used for the amplification of SSR loci in *Sorghum bicolor* and total number of accessions studied. SSR allele size ranges are those obtained in this study. The percentage of homogeneous accessions is given in parentheses for the global set (GS = the three subsets together) and each subset (Principal Component Score strategy: PCS, Logarithmic strategy: L, and Taxonomic strategy: T).

Locus	Repeat	Allele Size (bp)	GS	PCS	L	T
Multiplex set 1						
Sb4-15	(AG) <sub>16</sub>	118-134	562 (86)	208 (85)	196 (87)	158 (87)
Sb4-121	(AC) <sub>14</sub>	213-229	609 (85)	221 (80)	220 (86)	168 (91)
Sb4-32	(AG) <sub>15</sub>	167-235	575 (82)	201 (80)	219 (79)	155 (87)
Sb5-236	(AG) <sub>20</sub>	158-208	621 (76)	227 (70)	219 (81)	175 (75)
Sb6-342	(AC) <sub>25</sub>	268-294	555 (86)	211 (87)	201 (85)	143 (85)
Multiplex set 2						
Sb6-36	(AG) <sub>19</sub>	162-200	624 (58)	226 (61)	224 (55)	174 (59)
Sb1-1	(AG) <sub>16</sub>	242-300	616 (74)	223 (70)	219 (80)	174 (70)
Sb1-10	(AG) <sub>27</sub>	252-496	594 (75)	220 (69)	209 (81)	165 (77)
Sb5-256	(AG) <sub>8</sub>	166-182	537 (98)	203 (100)	198 (97)	136 (97)
Sb6-84	(AG) <sub>14</sub>	178-222	611 (80)	219 (74)	223 (87)	169 (77)
Multiplex set 3						
Sb6-42	(AG) <sub>26</sub>	166-212	614 (57)	224 (52)	219 (71)	171 (46)
Sb6-57	(AG) <sub>18</sub>	284-320	607 (72)	223 (68)	217 (84)	167 (61)
Sb6-34	[(AC)/(CG)] <sub>15</sub>	186-208	619 (86)	223 (83)	222 (89)	174 (86)
Sb5-206	(AC) <sub>13</sub> /(AG) <sub>20</sub>	104-156	605 (76)	219 (78)	212 (85)	174 (63)
Sb4-72	(AG) <sub>16</sub>	182-212	618 (83)	225 (83)	218 (87)	175 (78)

Table 2: Molecular diversity assessed with the allelic richness (No : total number of alleles) and with the genetic diversity ( $DI = 1 - \sum(p_i^2)$ , where  $p_i$  is the frequency of the  $i^{th}$  allele (Weir, 1989)) for the global set (GS = the three subsets together) and each subset (Principal Component Score strategy: PCS, Logarithmic strategy: L, and Taxonomic strategy: T) considering only homogeneous accessions.

Locus	GS	PCS	L	T
	No. (DI)	No. (DI)	No. (DI)	No. (DI)
Sb4-15	9 (0.80)	8 (0.81)	9 (0.76)	9 (0.77)
Sb4-121	9 (0.84)	9 (0.84)	9 (0.82)	8 (0.84)
Sb4-32	33 (0.91)	21 (0.91)	26 (0.89)	27 (0.91)
Sb5-236	24 (0.86)	19 (0.87)	19 (0.81)	19 (0.88)
Sb6-342	14 (0.79)	14 (0.80)	14 (0.77)	13 (0.79)
Sb6-36	20 (0.85)	17 (0.84)	19 (0.83)	18 (0.86)
Sb1-1	28 (0.86)	22 (0.83)	20 (0.88)	21 (0.85)
Sb1-10	31 (0.82)	26 (0.89)	20 (0.69)	23 (0.82)
Sb5-256	7 (0.27)	5 (0.30)	7 (0.23)	5 (0.28)
Sb6-84	23 (0.86)	21 (0.84)	20 (0.88)	16 (0.82)
Sb6-42	23 (0.93)	22 (0.93)	21 (0.92)	19 (0.90)
Sb6-57	16 (0.81)	12 (0.81)	15 (0.81)	11 (0.80)
Sb6-34	11 (0.74)	11 (0.76)	9 (0.71)	8 (0.74)
Sb5-206	26 (0.93)	25 (0.93)	24 (0.92)	23 (0.93)
Sb4-72	15 (0.71)	9 (0.72)	12 (0.66)	11 (0.75)
Total	289	241	244	231
Mean	19.3 (0.80)	16.1 (0.81)	16.3 (0.77)	15.4 (0.80)